

Jude, Nina [Hrsg.]; Klieme, Eckhard [Hrsg.]
PISA 2009 - Impulse für die Schul- und Unterrichtsforschung

Weinheim u.a. : Beltz 2013, 246 S. - (Zeitschrift für Pädagogik, Beiheft; 59)



Quellenangabe/ Reference:

Jude, Nina [Hrsg.]; Klieme, Eckhard [Hrsg.]: PISA 2009 - Impulse für die Schul- und Unterrichtsforschung. Weinheim u.a. : Beltz 2013, 246 S. - (Zeitschrift für Pädagogik, Beiheft; 59) - URN: urn:nbn:de:0111-opus-78209 - DOI: 10.25656/01:7820

<https://nbn-resolving.org/urn:nbn:de:0111-opus-78209>

<https://doi.org/10.25656/01:7820>

in Kooperation mit / in cooperation with:

BELTZ JUVENTA

<http://www.juventa.de>

Nutzungsbedingungen

Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Die Nutzung stellt keine Übertragung des Eigentumsrechts an diesem Dokument dar und gilt vorbehaltlich der folgenden Einschränkungen: Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen. Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use

We grant a non-exclusive, non-transferable, individual and limited right to using this document.

This document is solely intended for your personal, non-commercial use. Use of this document does not include any transfer of property rights and it is conditional to the following limitations: All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

Kontakt / Contact:

peDOCS
DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation
Informationszentrum (IZ) Bildung
E-Mail: pedocs@dipf.de
Internet: www.pedocs.de

59. Beiheft

April 2013

ZEITSCHRIFT FÜR PÄDAGOGIK

**PISA 2009 – Impulse für
die Schul- und Unterrichtsforschung**

BELTZ JUVENTA

Zeitschrift für Pädagogik · 59. Beiheft

PISA 2009 – Impulse für die Schul- und Unterrichts- forschung

Herausgegeben von
Nina Jude und Eckhard Klieme

BELTZ JUVENTA

Die in der Zeitschrift veröffentlichten Beiträge sind urheberrechtlich geschützt. Alle Rechte, insbesondere das der Übersetzung in fremde Sprachen, bleiben dem Beltz-Verlag vorbehalten.

Kein Teil dieser Zeitschrift darf ohne schriftliche Genehmigung des Verlages in irgendeiner Form – durch Fotokopie, Mikrofilm oder ein anderes Verfahren – reproduziert oder in eine von Maschinen, insbesondere Datenverarbeitungsanlagen, verwendbare Sprache übertragen werden. Auch die Rechte der Wiedergabe durch Vortrag, Funk- und Fernsehsendung, im Magnettonverfahren oder auf ähnlichem Wege bleiben vorbehalten. Fotokopien für den persönlichen oder sonstigen eigenen Gebrauch dürfen nur von einzelnen Beiträgen oder Teilen daraus als Einzelkopie hergestellt werden. Jede im Bereich eines gewerblichen Unternehmens hergestellte oder genutzte Kopie dient gewerblichen Zwecken gem. § 54 (2) UrhG und verpflichtet zur Gebührenzahlung an die VG Wort, Abteilung Wissenschaft, Goethestr. 49, 80336 München, bei der die einzelnen Zahlungsmodalitäten zu erfragen sind.

© 2013 Beltz Juventa · Weinheim und Basel

www.beltz.de · www.juventa.de

Herstellung: Lore Amann

Gesamtherstellung: Beltz Bad Langensalza GmbH, Bad Langensalza

E-Book

ISSN 0514-2717

Bestell-Nr. 443501

Inhaltsverzeichnis

Nina Jude/Eckhard Klieme
PISA 2009 – Impulse für die Schul- und Unterrichtsforschung.
Einleitung zum Beiheft 7

Schul- und Unterrichtsbedingungen

Brigitte Steinert/Jan Hochweber/Silke Hertel
Lesekompetenz und Lesefreude von Schülerinnen und Schülern und bildungs-
standardbezogene Kompetenzüberzeugungen und Lerngelegenheiten in Schule
und Unterricht 12

Silke Hertel/Simone Bruder/Nina Jude/Brigitte Steinert
Elternberatung an Schulen im Sekundarbereich. Schulische Rahmenbedingungen,
Beratungsangebote der Lehrkräfte und Nutzung von Beratung durch die Eltern .. 40

Timo Ehmke
Soziale Disparitäten im Lesen und in Mathematik innerhalb von Schulklassen .. 63

Schülerinnen und Schüler mit Migrationshintergrund

Aileen Edele/Petra Stanat/Susanne Radmann/Michael Segeritz
Kulturelle Identität und Lesekompetenz von Jugendlichen aus zugewanderten
Familien 84

*Robin Segerer/Alexandra Marx/Petra Stanat/Wolfgang Schneider/Thorsten Roick/
Peter Marx*
Determinanten der Lesekompetenz bei Jugendlichen nicht deutscher
Herkunftssprache. Zur Bedeutung der Spracherwerbsreihenfolge im Falle von
Mehrsprachigkeit 111

Trends und Veränderungen

Timo Ehmke/Eckhard Klieme/Petra Stanat

Veränderungen der Lesekompetenz von PISA 2000 nach PISA 2009. Die Rolle von Unterschieden in den Bildungswegen und in der Zusammensetzung der Schülerschaft	132
---	-----

Dorothea Mildner/Jan Hochweber/Andreas Frey

Vergleichende Analysen der Kompetenzen von Fünfzehnjährigen und Neuntklässlern in den deutschen PISA-Erhebungen 2003 bis 2009	151
---	-----

Linda Marie Bischof/Jan Hochweber/Johannes Hartig/Eckhard Klieme

Schulentwicklung im Verlauf eines Jahrzehnts – Erste Ergebnisse des PISA-Schulpanels	172
--	-----

Grundlagen und konzeptionelle Ansätze

Nina Jude/Johannes Hartig/Stefan Schipolowski/Katrin Böhme/

Petra Stanat

Definition und Messung von Lesekompetenz. PISA und die Bildungsstandards ..	200
---	-----

Eckhard Klieme/Svenja Vieluf

Schulische Bildung im internationalen Vergleich. Ein Rahmenmodell für Kontextanalysen in PISA	229
---	-----

Nina Jude/Eckhard Klieme

PISA 2009 – Impulse für die Schul- und Unterrichtsforschung

Einleitung

PISA, das Programme for International Student Assessment, ist im vergangenen Jahrzehnt zum Markenzeichen der empirischen Bildungsforschung geworden, aber auch zum Gegenstand heftigster Auseinandersetzungen um die empirische Beobachtung und politische Steuerung des Bildungswesens. Befürworter wie Kritiker tendieren gelegentlich dazu, die Bedeutung der Studie zu überschätzen, wenn sie in ihr einen gleichsam historischen Wendepunkt sehen. Im Kern ist PISA nichts anderes als eine alle drei Jahre stattfindende, repräsentativ und international vergleichend angelegte Schulleistungstudie, die Trenddaten zu Bedingungen, Prozessen und Ergebnissen schulischer Bildungsprozesse liefert. PISA setzt damit ein Paradigma der international vergleichenden Schulforschung fort, das seit Ende der 1960er Jahre in einer internationalen wissenschaftlichen Arbeitsgruppe, der International Association for the Evaluation of Student Achievement (IEA), entwickelt wurde. Die Erziehungswissenschaft ist gut beraten, solche Daten mit der nötigen theoretischen Einbettung und methodischen Sorgfalt zu nutzen, um dort, wo die Passung gegeben ist, Fragestellungen der Schulforschung, aber auch andere Themen wie etwa Einflüsse des Elternhauses auf die Entwicklung von Jugendlichen zu bearbeiten. Zudem bieten sich durch PISA völlig neue Möglichkeiten, kulturelle und schulstrukturelle Bedingungen sowie pädagogische Praktiken international vergleichend zu beschreiben und Veränderungen auf systemischer Ebene zu analysieren.

Im Dezember 2010 wurden die Befunde der bislang letzten Erhebung aus dem Jahr 2009 veröffentlicht. Damit liegen nun Daten aus insgesamt vier Erhebungszeitpunkten vor, die Veränderungen im Zeitraum von einem Jahrzehnt umfassen. Das vorliegende Beiheft zeigt anhand von national ausgerichteten, vertiefenden Analysen, welche erziehungswissenschaftlichen Forschungsfragen auf Basis dieser Daten bearbeitet und diskutiert werden können. Dabei steht das durch nationale Ergänzungen und Trendinformationen verstärkte Erkenntnispotenzial der PISA-Studien für substanzielle schulbezogene Forschung im Vordergrund.

Der populäre Blick auf PISA fokussiert die Kompetenzen von fünfzehnjährigen Schülerinnen und Schülern in den Bereichen Lesen, Mathematik und Naturwissenschaften im internationalen Vergleich. Aus erziehungswissenschaftlicher Sicht interessieren aber besonders Analysen der Schul- und Unterrichtsbedingungen, die auf umfangreichen – und in Deutschland durch die beteiligten Wissenschaftlerinnen und Wissenschaftler wesentlich erweiterten – Befragungen von Schülerinnen und Schülern, Eltern und Lehrpersonen beruhen. Diese Daten ermöglichen es, Rückschlüsse auf die Bedin-

gungen von Kompetenzentwicklung im deutschen Schulsystem zu ziehen und PISA-Daten für die Schul- und Unterrichtsforschung in Deutschland nutzbar zu machen. Dieses Beiheft versammelt Beiträge zu Schul- und Unterrichtsbedingungen, Kompetenzen von Schülerinnen und Schülern mit Migrationshintergrund, zu Trendanalysen für den Zeitraum zwischen 2000 und 2009 sowie zu den theoretischen Grundlagen und methodischen Ansätzen von PISA.

Fragestellungen der Schul- und Unterrichtsforschung lassen sich unter nationaler Perspektive auf Basis von PISA-Daten bearbeiten, weil die internationalen Fragebögen für Schülerinnen und Schüler, Eltern und Schulleitung in Deutschland ergänzt wurden und darüber hinaus eine nationale Zusatzstichprobe von zwei Klassen der 9. Jahrgangsstufe jeder teilnehmenden Schule getestet wurde. Erhoben wurden Kontextmerkmale auf Individual-, Klassen- und Schulebene, die Ansatzpunkte für die Analyse von Prozessen des Unterrichts und der Schule bieten. PISA 2009 setzte dabei einen Schwerpunkt auf Unterrichtsbedingungen und Unterrichtsqualität im Fach Deutsch sowie auf schulische und familiäre Strategien zur Förderung der Lesekompetenz. Darüber hinaus ermöglichen es Parallelisierungen in den nationalen Instrumenten, Merkmale von Schule und Unterricht aus Perspektive von Schülerinnen und Schülern sowie Lehrpersonen gleichermaßen zu erfassen und miteinander in Verbindung zu bringen. Bedingungsfaktoren für Kompetenzen und fachliches Interesse können so durch komplexe Angebot-Nutzungs-Modelle rekonstruiert werden.

Der Beitrag von Steinert, Hochweber und Hertel untersucht, inwieweit Kompetenzüberzeugungen von Lehrkräften und Praktiken zum Lesen im Deutschunterricht mit den *Lesekompetenzen und der Lesefreude von Jugendlichen* zusammenhängen. Es zeigt sich, dass die von Schülerinnen und Schülern einer Klasse geteilten Kompetenzüberzeugungen – auch bei Kontrolle der kognitiven und sozialen Klassenkomposition und der Schulart – für das Niveau der Lesekompetenz und der Lesefreude in der Klasse bedeutsam sind. Die Überzeugung der Schülerinnen und Schüler, Leseaufgaben gut bewältigen zu können, hängt wiederum mit den Überzeugungen des Lehrerkollegiums zusammen. Diese Befunde bestätigen, dass Aspekte der Schulkultur – also Werte, Normen und Einstellungen von Lehrenden und Lernenden – tatsächlich für das kognitive und affektive Lernen und die Lernergebnisse relevant sind.

Hertel, Bruder, Jude und Steinert gehen in ihrem Beitrag der Frage nach, wie die *schulischen Rahmenbedingungen für die Elternberatung* an Schulen im Sekundarbereich gestaltet sind. Neben Informationen über die Gestaltung der Beratungsangebote durch die Lehrkräfte zeigt der Artikel auch konkreten Handlungsbedarf auf, wenn es darum geht, Eltern mit niedrigem sozioökonomischen Hintergrund oder Migrationshintergrund in die Zusammenarbeit zwischen Schulen und Elternhaus einzubinden.

Soziale Disparitäten im Lesen und in Mathematik innerhalb von Schulklassen werden im Beitrag von Ehmke thematisiert. Neben dem Ausmaß der klassenspezifischen sozialen Disparitäten im Lesen und in der Mathematik und der Frage, ob klassenspezifische Herkunftseffekte eher domänenabhängig oder eher domänenübergreifend sind, wird auch analysiert, welche Bedeutung strukturellen Klassenmerkmalen und unterrichtsbezogenen Fördermaßnahmen zukommt.

Gerade durch PISA ist die wissenschaftliche und bildungspolitische Debatte um *migrationsbedingte Disparitäten im deutschen Bildungswesen* maßgeblich angestoßen worden. Im letzten Jahrzehnt lieferte PISA regelmäßig repräsentative Daten zum Zusammenhang zwischen Migrationshintergrund, Bildungslaufbahn und Lernergebnissen. Auch wenn sich die Kompetenzen der Schülerinnen und Schüler mit Migrationshintergrund im vergangenen Jahrzehnt im Durchschnitt signifikant verbessert haben, sind die Kompetenzunterschiede zu Schülerinnen und Schülern ohne Migrationshintergrund weiterhin gravierend. PISA 2009 knüpft an sozial- und erziehungswissenschaftliche Theorien zur Erklärung dieses Zusammenhangs an. In diesem Beiheft wird das Thema aus zwei unterschiedlichen Perspektiven aufgegriffen. Beide Beiträge liefern wichtige Hinweise auf Ansätze für Handlungsfelder in Schule und Gesellschaft.

Edele, Stanat, Radmann und Segeritz untersuchen die *kulturelle Identität und Lesekompetenz von Jugendlichen aus eingewanderten Familien*. Eine Erweiterung des PISA-Schülerfragebogens ermöglichte es, kulturelle Identitätsorientierungen und deren Zusammenhang mit dem Erwerb der deutschen Sprache erstmals repräsentativ zu untersuchen. Es zeigte sich, dass die meisten der in Deutschland lebenden Jugendlichen mit Migrationshintergrund stark mit dem Aufnahmekontext identifiziert sind. Wenn man sozialen Status und Familiensprache in Rechnung stellt, bleiben diese Jugendlichen in ihrer Deutsch-Lesekompetenz nicht hinter autochthonen Jugendlichen zurück (integrierte Gruppe) bzw. übertreffen diese sogar (assimilierte Gruppe). Eine Minderheit, die sich weder mit dem Aufnahme- noch mit dem Herkunftskontext stark identifiziert, ist jedoch besonders gefährdet, schwache schulbezogene Kompetenzen aufzuweisen.

Der Beitrag von Segerer, Marx, Stanat, Schneider, Roick und Marx diskutiert die Bedeutung von *Determinanten der Lesekompetenz bei Jugendlichen nichtdeutscher Herkunftssprache*. Dazu werden die Kompetenzen von mehrsprachigen und monolingualen leseschwachen Jugendlichen verglichen. Die Analysen beruhen auf einer ergänzenden Untersuchung, die an Schulen mit sehr hohem Migrantenanteil durchgeführt wurde. Neben dem PISA-Test zur Lesekompetenz wurden hierbei auch Messinstrumente eingesetzt, die den Wortschatz, grammatikalische Fähigkeiten und basale sprachliche Kompetenzen wie bspw. phonologische Bewusstheit und prosodische Sensitivität erfassten. Die Rückstände im Bereich des Leseverstehens von mehrsprachigen Jugendlichen werden unabhängig vom Zeitpunkt des Erwerbsbeginns der deutschen Sprache (Deutsch als Erstsprache oder Zweitsprache) primär mit eingeschränkten Fähigkeiten im Überschneidungsbereich von Grammatikkompetenz und Wortschatz erklärt.

Veränderungen über die Zeit können bei PISA im Sinne von Trends in den kognitiven Leistungen betrachtet werden. Aber auch Veränderungen in den Kontextbedingungen auf Schulebene sind von Interesse, wenn es darum geht, Fragen der Schulentwicklung als langfristigen Prozess zu diskutieren. Beide Perspektiven werden im vorliegenden Beiheft vertiefend in den Blick genommen.

Ehmke, Stanat und Klieme analysieren in ihrem Beitrag unter Nutzung eines ökonometrischen Ansatzes, welche Rolle die Unterschiede in den Bildungswegen und in der Zusammensetzung der Schülerschaft, u.a. in der Gruppe der Schülerinnen und Schüler mit Migrationshintergrund, für die *Veränderungen der Lesekompetenz* seit PISA 2000

spielen. Ein beträchtlicher Teil des für Deutschland ermittelten Anstiegs der Lesekompetenz lässt sich nach dieser Modellrechnung erklären durch die gestiegene Gymnasialquote und die Tatsache, dass späte Einschulung und Klassenwiederholungen seltener geworden sind, Fünfzehnjährige heutzutage also vermehrt in Jahrgangsstufe 10 und weniger in Jahrgangsstufe 9 oder darunter zu finden sind.

Übereinstimmende, obschon auf einem anderen methodischen Ansatz beruhende und auf Mathematik sowie Naturwissenschaften bezogene Erkenntnisse ermitteln Mildner, Hochweber und Frey, indem sie die *Veränderungen in den Kompetenzen von Fünfzehnjährigen mit den Veränderungen in den Kompetenzen der Neuntklässler* in den deutschen PISA-Erhebungen 2003 bis 2009 vergleichen. Das zunächst vielleicht überraschende Ergebnis lautet, dass alters- und jahrgangsstufenbezogene Trendanalysen auf Basis der PISA-Daten unterschiedlich ausfallen können. Während sich bei Fünfzehnjährigen signifikante Zugewinne zeigten, ist dies bei Neuntklässlern nicht der Fall. Eine Erklärung sehen auch Mildner, Hochweber und Frey darin, dass sich die Verteilung der Schülerinnen und Schüler auf Jahrgänge geändert hat. Vertiefende Auswertungen, die hier nicht mehr dokumentiert werden konnten, ergaben, dass die Veränderungen der Bildungswege zwischen 2000 und 2009 bei Jugendlichen mit Migrationshintergrund besonders stark ausgeprägt waren; die hier vorgelegten Modellrechnungen zur Erklärung der Leistungsentwicklung sind daher kompatibel mit der Aussage, dass signifikante Leistungsanstiege nur in der Gruppe der Migranten zu verzeichnen sind.

Bischof, Hochweber, Hartig und Klieme stellen in ihrem Beitrag erste Ergebnisse zur *Schulentwicklung im Verlauf eines Jahrzehnts* dar. Dazu nutzen sie Daten einer an PISA 2009 anschließenden, nationalen Panelstudie, in der Schulen, die an PISA 2000 teilgenommen hatten, im Jahr 2009 erneut getestet wurden. In ihren Analysen betrachten sie Faktoren der Schulorganisation, die Bedeutung von interner Evaluation als Schulentwicklungsmaßnahme sowie spezifischer Förderangebote auf Schulebene für die Unterschiede in der Entwicklung zwischen den Schulen. Erstmals wird für Deutschland empirisch nachgewiesen, dass interne Evaluierung sich sowohl auf den Leistungsstand als auch auf das Schulklima positiv auswirkt.

Abschließend betrachtet dieses Beiheft zwei konzeptionelle Themen. Der Beitrag von Jude, Hartig, Schipolowski, Böhme und Stanat zur *Definition und Messung von Lesekompetenz* diskutiert, inwieweit der international entwickelte Lesetest von PISA mit den in den letzten Jahren eingeführten, rein nationalen Tests der Studien zur Überprüfung der Bildungsstandards für den Mittleren Schulabschluss vergleichbar ist und wie die Kompetenzkonstrukte beider Instrumente zusammenhängen. Es zeigt sich, dass die in beiden Studien erfassten Lesekompetenzen zwar korrelieren und zu ähnlichen Aussagen über die Kompetenzverteilung bei den Schülerinnen und Schülern der neunten Jahrgangsstufe führen, aber doch konzeptionell und empirisch nicht als identische Konstrukte behandelt werden können. Dies impliziert, dass innerdeutsche Ländervergleiche, die auf den Bildungsstandards beruhen, nicht auf der PISA-Skala verankert werden können. Das internationale und das nationale Bildungsmonitoring werden also in Zukunft – jedenfalls im Bereich der Lesekompetenz – konzeptionell verwandte, aber numerisch nicht verrechenbare Befunde erbringen.

Ausblickend wird im Beitrag von Klieme und Vieluf ein *theoretisches Rahmenmodell für PISA* dargestellt, das zusammenfassend die Möglichkeiten und Grenzen der Studie aufzeigt und eine Perspektive für künftige PISA-Zyklen diskutiert. Das Modell beruht auf aktuellen Grundsatzdokumenten, welche die Autoren gemeinsam mit den internationalen Expertengruppen vorgelegt haben, die an den Fragebogen und den analytischen Auswertungsstrategien arbeiten. Diese an PISA unmittelbar beteiligten Wissenschaftlerinnen und Wissenschaftler sehen die Möglichkeiten, aber auch die Grenzen der Aussagefähigkeit der Studien durchaus kritisch. Die Dokumente belegen, dass PISA – zusätzlich zur Konzeption der Leistungstests – inzwischen eine differenzierte, an die erziehungswissenschaftliche und sonstige Bildungsforschung anschlussfähige theoretische Grundlage hat. Dadurch ergeben sich in der Zukunft neue Chancen, PISA-Daten für substantielle erziehungswissenschaftliche Forschung zu nutzen.

Anschrift der Autorin/des Autors

Dr. Nina Jude, Deutsches Institut für Internationale Pädagogische Forschung (DIPF),
Schloßstr. 29, 60486 Frankfurt am Main, Deutschland
E-Mail: jude@dipf.de

Prof. Dr. Eckhard Klieme, Deutsches Institut für Internationale Pädagogische Forschung
(DIPF), Schloßstraße 29, 60486 Frankfurt am Main, Deutschland
E-Mail: klieme@dipf.de

Schul- und Unterrichtsbedingungen

Brigitte Steinert/Jan Hochweber/Silke Hertel

Lesekompetenz und Lesefreude von Schülerinnen und Schülern und bildungsstandardbezogene Kompetenzüberzeugungen und Lerngelegenheiten in Schule und Unterricht

1. Einleitung

Die regelmäßige Erfassung und Beschreibung schulischer Rahmenbedingungen und Lerngelegenheiten gehören ebenso wie die regelmäßige Beobachtung und der internationale Vergleich ausgewählter Basiskompetenzen von 15-jährigen Schülerinnen und Schülern zum Programm der PISA-Studien. In Deutschland haben sich zwischen PISA 2000 und PISA 2009 nicht nur die Lesekompetenzen der Schülerinnen und Schüler verbessert, sondern auch einige Merkmale des schulischen Lernumfeldes wie das Schulklima und die Klassenführung (Hertel, Hochweber, Steinert & Klieme, 2010; Klieme, Artelt et al., 2010; Klieme, Jude, Baumert & Prenzel, 2010). Bei den meisten Merkmalen der schulischen Lernumgebung lagen die Einschätzungen der Befragten bei PISA 2009 innerhalb des OECD-Durchschnitts, bei einigen Merkmalen darunter.

Jedoch zeigten die länderspezifischen Analysen der OECD vielfach keinen direkten Zusammenhang zwischen den Lesekompetenzen der 15-Jährigen und den international erfassten Schulmerkmalen, insbesondere dann, wenn die soziale Herkunft der Schülerinnen und Schüler in den Schulen mit berücksichtigt wurde (OECD, 2011a). In Deutschland erstrecken sich die Analysen der schulischen Lernumgebung auf der Basis der nationalen Daten aus den bisherigen PISA-Zyklen weitgehend auf die Deskription von Schul- und Unterrichtsmerkmalen nach Schularten oder Bildungsgängen (Baumert, Trautwein & Artelt, 2003; Senkbeil et al., 2004; Senkbeil, Drechsel & Schöps, 2007; Weiß & Steinert, 2001) und die Typisierung von Schulen nach mehreren Schulmerkmalen (Senkbeil, 2006). Die im Rahmen der längsschnittlichen Erweiterungsstudie zu PISA 2003 durchgeführten Mehrebenenanalysen zeigten keinen Zusammenhang der verschiedenen Schulmerkmale mit dem Kompetenzzuwachs der Schülerinnen und Schüler von Klassenstufe neun auf zehn. Lediglich der durch aktive Lehrkräfte klassifizierte Schultypus zeigte einen Zusammenhang mit dem Kompetenzzuwachs in

Mathematik und Naturwissenschaften. Allerdings gingen in diese Typisierung Kompositionsmerkmale der kognitiven und sozialen Herkunft und allgemeine schulische Aktivitätsmerkmale ein, so dass die relative Bedeutung der jeweiligen Merkmale nicht unterschieden werden kann.

Demgegenüber fanden sich bei den Unterrichtsmerkmalen bedeutsame Effekte auf die Lernergebnisse der Schülerinnen und Schüler. Die dazu durchgeführten Analysen in der Mathematik in den national erweiterten PISA-2000- und PISA-2003-Studien erfolgten auch aufgrund vorausgegangener Studien (TIMSS-Video, Pythagoras-Projekt) in theoretisch wie empirisch elaborierter Weise (Baumert et al., 2010; Klieme, Pauli & Reusser, 2009; Klieme & Rakoczy, 2003; Klieme, Schümer & Knoll, 2001; Kunter et al., 2005). Dabei wurden drei Basisdimensionen der Unterrichtsqualität identifiziert: (1) klare, strukturierte Klassenführung, (2) unterstützendes Klassenklima, (3) kognitive Aktivierung. Mehrebenenanalysen mit diesen Dimensionen zeigten, dass die Leistungsunterschiede in Mathematik zwischen den Klassen auf die Unterrichtsmerkmale kognitive Aktivierung und Klassenführung zurückgeführt werden können. Als motivationales Ergebnismerkmal von Schule und Unterricht war das Interesse an Mathematik positiv mit dem Unterstützungsklima in den Klassen korreliert.

Angesichts des bisherigen Fokus auf die Bereiche Mathematik und Naturwissenschaften bei PISA ist es für die Schul- und Unterrichtseffektivitätsforschung wie für die Schulpraxis von Interesse, spezifische Lernbedingungen auf Schul- und auf Unterrichtsebene *im Bereich des Lesens* zu identifizieren, die Unterschiede in den Lesekompetenzen und der Lesemotivation der Schülerinnen und Schüler erklären und als Ausgangspunkt für die Gestaltung von Schule und Unterricht im Bereich des Lesens dienen können. Deshalb wird in diesem Beitrag untersucht, ob und inwieweit spezifische Kompetenzüberzeugungen und Praktiken zum Lesen im Deutschunterricht von Lehrkräften und von Schülerinnen und Schülern, die sich auf die Bildungsstandards im Fach Deutsch für den Bereich des Lesens beziehen (KMK, 2003, 2004), die Lesekompetenzen und die Lesefreude von Jugendlichen in der neunten Jahrgangsstufe vorhersagen können.

2. Theoretischer Hintergrund

Zur Erklärung der Unterschiede in der Lesekompetenz und der Lesefreude zwischen Schulen und Klassen wird zum einen auf formale Modelle und Konstrukte der Schul- und Unterrichtseffektivitätsforschung und zum anderen auf spezifische Konstrukte der Pädagogischen Psychologie zurückgegriffen. Formal ist das Schulwesen als Mehrebenensystem konzipiert, das die Handlungsebenen Schulsystem, Schule, Unterricht bzw. Klasse und die einzelnen Schülerinnen und Schüler unterscheidet und diesen jeweils spezifische Merkmale des Kontexts, des Inputs, der Prozesse und Ergebnisse zuordnet. Diese Prozess-Produkt-Modelle gehen davon aus, dass die Merkmale der höheren Ebenen primär für die Merkmale auf der jeweils niedrigeren Ebene bedeutsam sind, und dass deshalb Merkmale der Schule geringere Effekte auf die Lernergebnisse der Schü-

lerinnen und Schüler haben als Merkmale des Unterrichts. Metaanalysen zur Schul- und Unterrichtseffektivität berichten damit übereinstimmend die höchsten Effektstärken für die lernprozessnahen Schüleraktivitäten, gefolgt von lehrergesteuerten Unterrichtsmerkmalen und den Schulmerkmalen mit den niedrigsten Effektstärken (Fraser, Walberg, Welch & Hattie, 1987; Hattie, 2009; Scheerens & Bosker, 1997; Seidel & Shavelson, 2007; Wang, Haertel & Walberg, 1993). Dieses Befundmuster wird durch kognitive Modelle des Lehrens und Lernens gestützt, die die Lernaktivitäten der Schülerinnen und Schüler als proximale Merkmale für den Kompetenzerwerb konzipieren und die Unterrichts- bzw. Lehrervariablen als distale Merkmale (Seidel & Shavelson, 2007).

Auch wenn Unterrichts- und Lehrermerkmale mit spezifischem Bezug zur untersuchten Domäne konzipiert werden, dürften sie demnach wegen ihrer größeren Distanz zu den Lernprozessen kleinere Effektstärken auf die spezifischen Lernaktivitäten und Lernergebnisse der Schülerinnen und Schüler aufweisen. Für Schulmerkmale ist dagegen theoretisch kaum von einem direkten Effekt auf die Kompetenzentwicklung der Schülerinnen und Schüler auszugehen, doch könnten die gemeinsam geteilten Überzeugungen und Praktiken der Lehrkräfte auf Schulebene als organisatorischer und sozialer Handlungskontext über die entsprechenden Überzeugungen und Praktiken im Unterricht die Schülerleistungen indirekt begünstigen, wenn sie diese in der Organisation von Lernprozessen über Jahrgangsstufen, Klassen und Fächer hinweg in einen systematischen Zusammenhang stellen. Angesichts der tatsächlichen Unterschiede zwischen Klassen und der professionellen Autonomie der Lehrkräfte im Unterricht wäre ein moderater Einfluss der Lehrereinschätzungen und schulischen Praktiken auf den Unterricht zu erwarten (Hochweber, Steinert & Klieme, 2012; Klieme, Steinert & Hochweber, 2010; Rolff, 2007; Spiess, 2004; Steinert, Hartig & Klieme, 2008; Steinert et al., 2006).

Diese Annahmen werden auch von Modellen des Prozess-Produkt-Paradigmas sowie Modellen des zielerreichenden Lernens bzw. Unterrichts und teilweise von Befunden der Schulleistungsforschung gestützt. Danach sind klare Ziele, hohe Leistungserwartungen und der Fokus auf das Lernen der Schülerinnen und Schüler unter den Lehrkräften einer Schule Voraussetzung für eine lernförderliche Unterrichtskultur und bessere Lernergebnisse. Auch das Angebot an zielgerichteten Lerngelegenheiten soll zu höheren Leistungen beitragen (vgl. Bloom, 1976; Carroll, 1963; Creemers & Kyriakides, 2008; Ditton, 2000, 2007; Edmonds, 1979; Fend, 2006; Helmke & Weinert, 1997; Rosenshine, 1979; Sammons, Hillman & Mortimore, 1995; Scheerens & Bosker, 1997; Slavin, 1996; Stringfield, 1994; Teddlie & Reynolds, 2000). Die empirischen Befunde sind allerdings nicht eindeutig: In den nach Domänen, motivationalen und prozessualen Ergebniskriterien differenzierten Metaanalysen von Seidel und Shavelson (2007) erwiesen sich Lehrerüberzeugungen und hohe Leistungserwartungen – sogar als distale Merkmale – als prädiktiv für die kognitiven und motivationalen Lernergebnisse im Lesen; Lernzeiten und Lerngelegenheiten waren für die Lesekompetenzen prädiktiv. Andere Studien zeigten jedoch, dass die Effekte der Lernzeiten und des zielerreichenden Unterrichts auf die Lernergebnisse verschwinden, wenn standardisierte Leistungstests eingesetzt werden, da Lernzeiten und Leistungserwartungen mit den Lernvoraussetzungen der Schülerinnen und Schüler konfundiert sind (vgl. Gruehn, 2000). Analysen der

Daten zum Deutschunterricht auf der Basis der DESI-Studie haben zudem gezeigt, dass Lehrkräfte auf schwache Lernausgangslagen bzw. ein niedriges Leistungsniveau in ihren Klassen adaptiv reagieren, indem sie verstärkt vielfältige Unterrichtsmethoden und Übungsaufgaben einsetzen (Klieme et al., 2008). Solch ein remediales Vorgehen kann sich in Schulleistungsuntersuchungen als fehlender oder möglicherweise sogar negativer Zusammenhang zwischen den Lesegelegenheiten im Unterricht und den Lesekompetenzen darstellen (Klieme, Steinert & Hochweber, 2010).

In der nationalen Erweiterung der PISA-2009-Studie wurden die Lerngelegenheiten und Leistungserwartungen der Lehrkräfte und Schülerinnen und Schüler nach fach- bzw. domänenspezifischen Kriterien erfasst, um einen sachlichen Bezugsrahmen für die Einschätzungen der Befragten zu schaffen. Dazu wurden die Fragen zu den im Unterricht verwendeten Leseaufgaben und zu den Kompetenzüberzeugungen im Hinblick auf die Aufgabenanforderungen formuliert, die die KMK in den Bildungsstandards für den Bereich des Lesens im Fach Deutsch in der Sekundarstufe I erlassen hat (KMK, 2003, 2004) (zur Operationalisierung siehe Abschnitt 4.2).

Die so erfassten Kompetenzüberzeugungen zum Lesen können dabei auch mit Bezug auf das Konstrukt der Selbstwirksamkeit von Bandura (1977, 1997) konzipiert werden, indem sie die Zuversicht von Schülerinnen und Schülern sowie von Lehrkräften erfassen, dass die Schülerinnen und Schüler die in den Bildungsstandards formulierten Leseanforderungen erfolgreich bewältigen können. Theoretisch wird davon ausgegangen, dass Kompetenzüberzeugungen zum einen auf früheren (Lese-)Erfahrungen und Erfolgen in ähnlichen Leistungssituationen beruhen, zum anderen auf der Überzeugung, auch in neuen Situationen über die für das angestrebte Ziel (Leseverständnis) relevanten Handlungsmittel (Fähigkeit, Anstrengung, einflussreiche Lehrkräfte, Glück) zu verfügen und diese erfolgreich einsetzen zu können (Köller & Möller, 2010). Auch wenn diese Mittel in der national erweiterten PISA-2009-Studie nicht erhoben wurden, wird mit den auf das Lesen bezogenen Kompetenzüberzeugungen ein Merkmal erhoben, für das eine hohe Prädiktionskraft für Unterschiede in Lesekompetenz und Lesemotivation anzunehmen ist. Kompetenzüberzeugungen haben als Konstrukt auch eine hohe praktische Relevanz, weil sie sich im Unterricht durch Lehrkräfte gut fördern lassen. Dies kann beispielsweise dadurch geschehen, dass Schülerinnen und Schüler z.B. ihre Verantwortlichkeit für Leseerfolge und -fortschritte erleben, effektive Lesestrategien nutzen, Erfahrungen mit vielfältigen und unterschiedlich schwierigen Texten sammeln und konstruktive, aufgabenspezifische Rückmeldungen erhalten. Denn positive Erfahrungen der Lesenden in der Bewältigung langer und schwieriger Texte stützen die Ausdauer und Bereitschaft zu lesen und die Vielfalt und Häufigkeit des Lesens (Artelt et al., 2005; McElvany & Becker, 2010). Die Leseforschung geht überdies von einer wechselseitigen Beziehung zwischen Lesemotivation und Lesekompetenz im Sinne eines sich selbst verstärkenden Mechanismus aus (Matthäus-Effekt), nach dem Leseerfahrung und Leseerfolge zu einer weiteren Steigerung von Lesekompetenz und Lesemotivation führen (Artelt, Naumann & Schneider, 2010).

Im Folgenden werden die bildungsstandardbezogenen Kompetenzüberzeugungen und Lerngelegenheiten im Bereich des Lesens im Fach Deutsch auf der Basis der na-

tionalen Erweiterungsstichprobe von PISA 2009 mit zwei intakten neunten Klassen analysiert. Dabei wird ein Angebots-Nutzungs-Modell im Sinne von Fend (2008) zugrunde gelegt. Auf der Angebotsseite und Schulebene werden a) die gemeinsam geteilten Kompetenzüberzeugungen der Deutsch-Lehrkräfte hinsichtlich der Bewältigung der in den Bildungsstandards formulierten Leseanforderungen durch die Schülerinnen und Schüler, b) die gemeinsam geteilten Kompetenzziele im Lesen, c) die eingesetzten Leseaufgaben im Deutschunterricht und d) die Absprache von Leistungsanforderungen verwendet. Auf der Nachfrageseite und Klassenebene werden a) die gemeinsam geteilten Kompetenzüberzeugungen der Schülerinnen und Schüler, die in den Bildungsstandards formulierten Leseanforderungen bewältigen zu können, und b) die Nutzung der entsprechenden Leseaufgaben im Unterricht herangezogen. Die Merkmale der Schulebene sollten dabei für die Unterrichtsmerkmale, die Unterrichtsmerkmale wiederum für Lesekompetenz und Lesefreude Vorhersagekraft besitzen. Diese Modellierung trägt organisationstheoretischen und pädagogischen Überlegungen Rechnung, nach denen Einzelschulen die pädagogischen Handlungs- und Entwicklungseinheiten sind, in denen die Lehrkräfte über soziale Interaktionen sowie soziale Vergleichs- und Bewertungsprozesse gemeinsam geteilte, professionelle Normen und Praktiken etablieren und damit die schulische Lernumgebung steuern, welche Schülerinnen und Schüler im Unterricht nutzen (Chrispeels & Harris, 2006; Dreeben & Barr, 1988; Fend, 2006, 2008).

Konkret wird für die nachfolgenden Analysen davon ausgegangen, dass zwischen den gemeinsam geteilten Kompetenzüberzeugungen der Lehrkräfte auf Schulebene und den entsprechenden Überzeugungen der Schülerinnen und Schüler auf Klassenebene ein moderater Zusammenhang besteht, ebenso zwischen den Lerngelegenheiten im Lesen auf Schulebene und der Nutzung von Leseaufgaben im Deutschunterricht. Die Absprache von Leistungsanforderungen als Merkmal der Koordination und Kooperation dürfte nach Befunden aus der DESI-Studie vermutlich mit den Lerngelegenheiten im Unterricht korrelieren (Klieme, Steinert & Hochweber, 2010). Die Effekte der Merkmale auf der Schul- und Unterrichtsebene auf Unterricht bzw. Lernergebnisse der Schülerinnen und Schüler dürften sich ferner reduzieren, wenn die soziale Komposition der Schulklassen und die Schulart berücksichtigt werden. Die Bedeutsamkeit der individuellen Ausgangslagen der Schülerinnen und Schüler, der sozialen Komposition der Schulklassen und der Schularten als differenzielle Entwicklungsmilieus für den Kompetenzerwerb wurde in nationalen (und internationalen) Schulleistungsstudien vielfach gezeigt (Baumert & Schümer, 2001, 2002; Ehmke & Baumert, 2007; Ehmke, Hohensee, Heidemeier & Prenzel, 2004; Ehmke & Jude, 2010; OECD, 2001, 2010; Stanat, Rauch & Segeritz, 2010).

3. Fragestellungen

In diesem Beitrag wird anhand der national erweiterten Stichprobe von PISA 2009 der Frage nachgegangen, ob und inwieweit die von Lehrkräften bzw. von Schülerinnen und Schülern geteilten, auf die Bildungsstandards im Bereich des Lesens bezogenen Kom-

petenzüberzeugungen und Praktiken zum Lesen im Deutschunterricht die Lesekompetenz und Lesefreude der Schülerinnen und Schüler vorhersagen können. Dazu werden die in Abschnitt 2 formulierten Zusammenhänge zwischen Schulmerkmalen, Unterrichtsmerkmalen und Schülerkompetenzen in Bezug auf die Überzeugungen und Praktiken von Lehrkräften und Schülerinnen und Schülern schrittweise spezifiziert.

Die Lesekompetenz ist eine zentrale, fächerübergreifende Schlüsselqualifikation, die nicht ausschließlich in der Schule und dort nicht nur im Deutschunterricht erworben wird. Dennoch ist davon auszugehen, dass institutionell gestaltbare Merkmale, hier die Kompetenzüberzeugungen und Praktiken von Lehrkräften und Schülerinnen und Schülern, auf Schul- und Unterrichtsebene für Unterschiede in der Lesekompetenz und Lesefreude von Schülerinnen und Schülern direkt oder indirekt bedeutsam sind.

Diese allgemeine Annahme impliziert, dass die beiden Zielkriterien nicht nur individuell zwischen Schülerinnen und Schülern variieren, sondern auch institutionell zwischen Schulen und zwischen Schulklassen. Befunde für diese Annahme liegen aus früheren PISA-Studien und vergleichbaren Schulleistungsstudien vor (Baumert & Schümer, 2001; Baumert et al., 2003; Klieme, Steinert & Hochweber, 2010; Prenzel, Senkbeil & Drechsel, 2004). In einem ersten Analyseschritt wird daher geprüft, inwieweit die Lesekompetenz und die Lesefreude auf Schul-, Klassen- und Schülerebene variieren.

In einem zweiten Analyseschritt wird der Frage nachgegangen, ob und inwieweit die bildungsstandardbezogenen Kompetenzüberzeugungen und Praktiken von Lehrkräften und von Schülerinnen und Schülern auf Schul- und Unterrichtsebene in einem Zusammenhang stehen. Konkret wird untersucht, ob die auf Schulebene gemeinsam geteilten Kompetenzüberzeugungen, Kompetenzziele, angebotenen Leselernegelegenheiten und Absprachen von Leistungsanforderungen unter den Lehrkräften des Fachkollegiums Deutsch mit den entsprechenden Kompetenzüberzeugungen und den von den Schülerinnen und Schülern wahrgenommenen Leselernegelegenheiten im Deutschunterricht auf Klassenebene – wie im vorigen Abschnitt 2 postuliert – zusammenhängen.

Im dritten Analyseschritt wird schließlich untersucht, ob und inwieweit die gemeinsam geteilten bildungsstandardbezogenen Kompetenzüberzeugungen und Praktiken der Lehrkräfte und der Schülerinnen und Schüler auf der Schul- bzw. auf der Unterrichtsebene – auch unter Berücksichtigung der individuellen und kompositionellen Ausgangslagen sowie der Schulart – für Unterschiede in der Lesekompetenz und Lesefreude bedeutsam sind. Nach den in Abschnitt 2 ausgeführten Überlegungen, dass von einer Variation des Unterrichts innerhalb von Schulen aus pädagogischen und strukturellen Gründen (adaptiver Umgang mit unterschiedlichen Lernvoraussetzungen der Schülerinnen und Schüler in den Klassen, professionelle Autonomie der Lehrkräfte) auszugehen ist, dürfte kein direkter Zusammenhang zwischen den Schulmerkmalen und den Lesekompetenzen und der Lesefreude der Schülerinnen und Schüler bestehen, wenn die entsprechenden Kompetenzüberzeugungen und wahrgenommenen Lerngelegenheiten seitens der Schülerinnen und Schüler mit in die Analyse einbezogen und wenn die Effekte der individuellen und kompositionellen Hintergrundmerkmale sowie der Schulart kontrolliert werden (Ditton & Kreckler, 1995; Klieme, Steinert & Hochweber, 2010).

Ferner wird davon ausgegangen, dass den lesebezogenen Kompetenzüberzeugungen der Schülerinnen und Schüler positive Leseerfahrungen und die Überzeugung vorausgehen, auch über die für die Bewältigung neuer Leseaufgaben notwendigen Strategien zu verfügen, so dass ein positiver Zusammenhang zwischen den lesebezogenen Kompetenzüberzeugungen der Schülerinnen und Schüler und den Lesekompetenzen und der Lesefreude angenommen wird. Das Angebot an Lerngelegenheiten mit bildungsstandardbezogenen Leseaufgaben im Unterricht könnte mit der Lesekompetenz und Lesemotivation in einem positiven Zusammenhang stehen. Bei einem überwiegend remedialen Einsatz von Leseaufgaben im Unterricht ist jedoch ein negativer Zusammenhang wahrscheinlich.

Schließlich wird angenommen, dass Unterschiede in der Lesekompetenz und Lesefreude nicht nur durch Unterschiede in den Kompetenzüberzeugungen und Praktiken der Schülerinnen und Schüler erklärt werden können, sondern auch durch kompositionelle Merkmale und die Schulart. Die Schulart dürfte insbesondere für die Lesekompetenz, weniger für die Lesefreude von Bedeutung sein.

4. Datenbasis und methodisches Vorgehen

4.1 Stichprobe und Design

Die folgenden Analysen basieren auf den Daten der nationalen Erweiterungsstichprobe der PISA-2009-Studie. Diese geht an drei Stellen über das internationale Design von PISA hinaus (Hertel et al., 2010): (1) Die Stichprobe der Schülerinnen und Schüler wurde auf zwei komplette Klassen der neunten Jahrgangsstufe an jeder PISA-2009-Schule erweitert. (2) Es wurde ein nationaler Fragebogen für Schülerinnen und Schüler mit Fokus auf das Lesen und Lernen im Deutschunterricht eingesetzt. (3) Mit einem weiteren nationalen Fragebogen wurden die Überzeugungen und die Praktiken in Bezug auf das Lesen im Deutschunterricht der Lehrerinnen und Lehrer des Fachkollegiums Deutsch an den PISA-2009-Schulen erfasst.

Datenbasis dieses Beitrags sind die Testergebnisse zur Lesekompetenz sowie die Fragebogenangaben zur Lesemotivation und zu den Lernbedingungen im Deutschunterricht von 9193 Schülerinnen und Schülern in 382 neunten Schulklassen an 197 Schulen, die an der PISA-2009-Studie teilgenommen haben. Außerdem wurden die Angaben von 2113 Lehrpersonen, die das Fach Deutsch an den PISA-2009-Schulen unterrichten, zu ihren Überzeugungen und Praktiken in die Analysen einbezogen.

4.2 Instrumente

In den nachfolgenden Drei-Ebenen-Analysen werden die Lesekompetenz und die Lesefreude der Schülerinnen und Schüler der neunten Jahrgangsstufe als Zielkriterien schulischen Lernens untersucht. Prädiktoren zur Erklärung der Unterschiede in den kogni-

tiven und motivationalen Lernergebnissen sind auf Schulebene neben der Schulart vier Merkmale, die die geteilte Wahrnehmung der Lehrkräfte des Fachkollegiums Deutsch im Hinblick auf die Schülerinnen und Schüler der neunten Jahrgangsstufe widerspiegeln: (1) die Kompetenzüberzeugungen der Deutsch-Lehrkräfte hinsichtlich der Bewältigung der in den Bildungsstandards formulierten Leseanforderungen durch die Schülerinnen und Schüler, (2) die auf die Bildungsstandards bezogenen Kompetenzziele im Bereich des Lesens, (3) die Übungsintensität bzw. Häufigkeit von auf die Bildungsstandards bezogenen Leseaufgaben im Deutschunterricht, (4) die Absprache von Zielen und Leistungsanforderungen.

Auf Individual- und Klassenebene gehen die individuellen und sozial geteilten, auf die Bildungsstandards im Bereich des Lesens bezogenen Kompetenzüberzeugungen der Schülerinnen und Schüler und deren Wahrnehmung der Übungsintensität bzw. Häufigkeit entsprechender Leseaufgaben im Deutschunterricht in die Analyse ein. Kontrollvariablen für den sozialen und kognitiven Hintergrund der Schülerinnen und Schüler sind das Geschlecht, der Migrationshintergrund, der sozioökonomische Status, die kognitiven Grundfähigkeiten sowie deren Komposition auf Klassenebene.

Um den Bezug der Kompetenzüberzeugungen und Leseaufgaben im Deutschunterricht zu den Bildungsstandards herzustellen, wurden 20 verschiedene Kompetenzaspekte aus den Bildungsstandards für den Bereich Lesen im Fach Deutsch für den Mittleren Schulabschluss ausgewählt (KMK, 2003). Den Lehrkräften und den Schülerinnen und Schülern wurden jeweils die gleichen Items vorgelegt. Während die Schülerinnen und Schüler nach ihren eigenen Kompetenzüberzeugungen gefragt wurden, wurde den Lehrpersonen die Frage gestellt, welche Kompetenzüberzeugungen sie im Hinblick auf die Schülerinnen und Schüler der neunten Jahrgangsstufe haben (siehe Abschnitt 4.2, Schulmerkmale: Kompetenzüberzeugungen der Lehrpersonen und angebotene Leseaufgaben im Deutschunterricht, und vgl. die Dokumentation der Skalen bei Hertel et al., in Vorb.).

Die verschiedenen Kompetenzaspekte des Lesens wurden einer Konfirmatorischen Faktorenanalyse (CFA) unterzogen, um zentrale Dimensionen des Lesens zu beschreiben und die Äquivalenz der ermittelten Dimensionen aus Lehrer- und aus Schülersicht zu prüfen. Sowohl in den Wahrnehmungen der Lehrkräfte als auch in den Wahrnehmungen der Schülerinnen und Schüler zeigten die verschiedenen Kompetenzaspekte des Lesens die gleiche Dimensionalität: Reflektieren und Bewerten von Texten (9 Items), textbezogenes Kombinieren und Interpretieren (5 Items), Umgang mit diskontinuierlichen Texten (6 Items) (siehe Abschnitt 4.2, Schulmerkmale: Kompetenzüberzeugungen der Lehrpersonen und angebotene Leseaufgaben im Deutschunterricht, und Abschnitt 4.2, Unterrichtsmerkmale: Kompetenzüberzeugungen der Schülerinnen und Schüler und wahrgenommene Leseaufgaben im Deutschunterricht).

Im vorliegenden Beitrag werden Ergebnisse zu den bildungsstandardbezogenen Kompetenzüberzeugungen und Praktiken der Lehrkräfte und Schülerinnen und Schüler anhand der Dimension Reflektieren und Bewerten von Texten dargestellt (siehe Abschnitt 5). Diese enthält sehr komplexe Anforderungen, bei der Schülerinnen und Schüler in Deutschland in den früheren PISA-Studien relative Schwächen zeigten. Zu den

beiden anderen Dimensionen werden noch analoge Analysen durchgeführt, deren Darstellung allerdings den Rahmen dieses Beitrags überschreiten würde.

Schulmerkmale: Kompetenzüberzeugungen der Lehrpersonen und angebotene Leseaufgaben im Deutschunterricht

Die *auf die Bildungsstandards bezogenen Kompetenzüberzeugungen im Lesen* wurden über die Frage erfasst, wie die Lehrpersonen des Fachkollegiums Deutsch die Leistungen der Schülerinnen und Schüler ihrer Schule am Ende der neunten Jahrgangsstufe im Hinblick auf die Bewältigung der in den Bildungsstandards für den Bereich des Lesens im Fach Deutsch formulierten Leseaufgaben einschätzen. In einem 4-stufigen Antwortformat sollten die Lehrpersonen angeben, wie viele Schülerinnen und Schüler der neunten Jahrgangsstufe an der jeweiligen Schule zum Ende des Schuljahres einzelne Leseaufgaben bewältigen können (unter 70%, 71%-80%, 81%-90%, über 90%). Ein Beispielitem für die Skala „Reflektieren und Bewerten von Texten“ lautet: „Entscheiden, ob die Informationen in einem Text Tatsachen sind oder eine Meinung wiedergeben.“ Die drei Skalen aus der Befragung der Lehrkräfte des Fachkollegiums Deutsch – Reflektieren und Bewerten von Texten, textbezogenes Kombinieren und Interpretieren, Umgang mit diskontinuierlichen Texten – weisen jeweils eine sehr hohe interne Konsistenz auf (Cronbachs $\alpha = .92$). Das mit einer Konfirmatorischen Faktorenanalyse (CFA) geprüfte Modell mit drei Dimensionen für die bildungsstandardbezogenen Kompetenzerwartungen aus Sicht der Lehrpersonen weist eine akzeptable Passung zu den Daten auf ($\text{Chi}^2 = 1692.908$, $df = 167$, $p < 0.001$; RMSEA = .070, CFI = .97).¹

Die Bedeutung von *Kompetenzziele im Lesen im Deutschunterricht* wurde über die Frage erfasst, wie wichtig es der Lehrkraft ist, dass Schülerinnen und Schüler an ihrer Schule am Ende der neunten Jahrgangsstufe die bildungsstandardbezogenen Leseaufgaben bewältigen können. In einem 4-stufigen Antwortformat konnten die Lehrpersonen die Wichtigkeit von „überhaupt nicht wichtig“ bis „sehr wichtig“ bewerten. Die drei Skalen zur Wichtigkeit von Lesekompetenzziele im Deutschunterricht bei den Lehrpersonen weisen eine hohe bis sehr hohe interne Konsistenz auf: Reflektieren und Bewerten von Texten ($\alpha = .86$), textbezogenes Kombinieren und Interpretieren ($\alpha = .80$), Umgang mit diskontinuierlichen Texten ($\alpha = .91$). Das in einer CFA geprüfte Modell mit den drei Dimensionen für die Kompetenzziele aus Sicht der Lehrpersonen weist eine akzeptable Passung zu den Daten auf ($\text{Chi}^2 = 1591.744$, $df = 167$, $p < 0.001$; RMSEA = .066, CFI = .96).

Die *Häufigkeit bzw. Übungsintensität von bildungsstandardbezogenen Leseaufgaben im Deutschunterricht* wurde über die Frage erfasst, wie häufig die Lehrpersonen

1 Die Beurteilung der Modellanpassung erfolgte primär auf Grundlage von Fitindizes, da der Chi^2 -Modelltest sehr sensitiv auf die Größe der Stichprobe reagiert (Bentler & Bonett, 1980). Basierend auf Empfehlungen in der einschlägigen Literatur (z.B. Brown, 2006; Browne & Cudeck, 1993; Hu & Bentler, 1999) können Werte des CFI $\geq .90$ und $\geq .95$ als akzeptabel bzw. gut eingestuft werden. Werte des RMSEA $\leq .05$ werden i.A. als gut, Werte $\leq .08$ als akzeptabel, Werte $> .10$ als ungenügend bewertet.

den Schülerinnen und Schülern der neunten Jahrgangsstufe die einzelnen Leseaufgaben stellen. In einem 4-stufigen Antwortformat konnten die Lehrpersonen die Häufigkeit der gestellten Leseaufgaben mit „nie oder fast nie“, „in einigen Stunden“, „in den meisten Stunden“, „in jeder Stunde“ beantworten. Die drei Skalen zur Häufigkeit von Leseaufgaben im Deutschunterricht weisen eine hohe interne Konsistenz auf: Reflektieren und Bewerten von Texten ($\alpha = .87$), textbezogenes Kombinieren und Interpretieren ($\alpha = .87$), Umgang mit diskontinuierlichen Texten ($\alpha = .89$). Das dreidimensionale Modell für die Häufigkeit der Leseaufgaben im Deutschunterricht aus Sicht der Lehrpersonen weist eine akzeptable Passung zu den Daten auf ($\chi^2 = 1783.402$, $df = 167$, $p < 0.001$; RMSEA = .070, CFI = .96).

Außerdem wurden die Lehrpersonen auch nach der *Absprache von Zielen und Leistungsanforderungen* gefragt, um den Abstimmungs- und Koordinationsbedarf im Fachkollegium Deutsch zu ermitteln, der bei der Vorgabe von Bildungsstandards besteht. Sie konnten die Frage auf einer 4-stufigen Antwortskala von „trifft überhaupt nicht zu“ bis „trifft genau zu“ beantworten. Ein Beispielitem für diese Skala mit fünf Items und einer zufriedenstellenden internen Konsistenz ($\alpha = .78$) lautet: „Die Leistungsanforderungen im Fach Deutsch sprechen wir auch außerhalb von Vergleichsarbeiten ab.“

Die Skalen aus der Lehrerbefragung des Fachkollegiums Deutsch wurden in den folgenden Drei-Ebenen-Modellen auf Schulebene aggregiert und als Indikatoren für die innerhalb der Schule gemeinsam geteilten, auf die Bildungsstandards bezogenen Kompetenzüberzeugungen und Praktiken der Deutsch-Lehrpersonen verwendet.

Unterrichtsmerkmale: Kompetenzüberzeugungen der Schülerinnen und Schüler und wahrgenommene Leseaufgaben im Deutschunterricht

Die *auf die Bildungsstandards bezogenen Kompetenzüberzeugungen im Lesen* der Schülerinnen und Schüler wurden über die Frage erfasst, wie gut die Schülerin bzw. der Schüler der neunten Jahrgangsstufe die einzelnen Leseaufgaben lösen kann. Das 4-stufige Antwortformat sah folgende Antwortalternativen vor: „gar nicht“, „weniger gut“, „gut“ und „sehr gut“. Die Formulierung der Items für die einzelnen Leseaspekte hatte in der Schülerbefragung den gleichen Wortlaut wie in der Lehrerbefragung. Die drei Skalen zu den Kompetenzerwartungen der Schülerinnen und Schüler weisen eine zufriedenstellende interne Konsistenz auf – Reflektieren und Bewerten von Texten ($\alpha = .81$), textbezogenes Kombinieren und Interpretieren ($\alpha = .74$), Umgang mit diskontinuierlichen Texten ($\alpha = .80$). In der CFA weist das dreidimensionale Modell für die Kompetenzerwartungen der Schülerinnen und Schüler eine noch akzeptable Passung zu den Daten auf ($\chi^2 = 4179.136$, $df = 167$, $p < 0.001$; RMSEA = .053, CFI = .90).

Die Wahrnehmung der *Übungsintensität bzw. Häufigkeit der auf die Bildungsstandards bezogenen Leseaufgaben im Deutschunterricht* durch die Schülerinnen und Schüler wurde über die Frage erfasst, wie oft die Schülerin bzw. der Schüler die folgenden Aufgabenstellungen im Deutsch-Unterricht bzw. bei den Deutsch-Hausaufgaben bearbeitet. Die Frage konnte in einem 4-stufigen Antwortformat mit „nie oder fast nie“, „in einigen Stunden“, „in den meisten Stunden“, „in jeder Stunde“ beantwortet werden. Die

drei Skalen zur Häufigkeit von Leseaufgaben im Deutschunterricht weisen eine hohe interne Konsistenz auf: Reflektieren und Bewerten von Texten ($\alpha = .84$), textbezogenes Kombinieren und Interpretieren ($\alpha = .81$), Umgang mit diskontinuierlichen Texten ($\alpha = .87$). Das dreidimensionale Modell zeigt eine noch akzeptable Passung zu den Daten ($\chi^2 = 9146.690$, $df = 167$, $p < 0.001$; RMSEA = .080, CFI = .89).

Die Skalen aus der Schülerbefragung wurden in den folgenden Drei-Ebenen-Modellen als individuelle Wahrnehmungen und – auf Klassenebene aggregiert – als Indikatoren für die innerhalb der Klasse geteilten bildungsstandardbezogenen Kompetenzüberzeugungen und die Nutzung von Leseangeboten im Deutschunterricht verwendet.

Soziobiografische Hintergrundmerkmale und Komposition

Als Kovariaten werden im Folgenden vier soziobiografische Merkmale der Schülerinnen und Schüler, die Komposition dieser Merkmale auf der Ebene der Schulklassen und die Schulart als Indikator für das institutionelle Lernmilieu berücksichtigt.

Sozioökonomischer Status: Der sozioökonomische Status der Schülerfamilie wurde mithilfe des internationalen sozioökonomischen Indexes (*international socio-economic index*; ISEI) nach Ganzeboom, de Graaf, Treiman & de Leeuw (1992) erfasst. Als Indikator kam der höchste ISEI der Mutter bzw. des Vaters der Schülerinnen und Schüler (HISEI) zum Einsatz.

Allgemeine kognitive Fähigkeit: Die allgemeine kognitive Fähigkeit der Schülerinnen und Schüler wurde mittels der zweiten nicht-verbalen Skala des Kognitiven Fähigkeitstests (KFT; Heller & Perleth, 2000) erfasst, in der figurale Analogien zum Einsatz kommen.

Migrationshintergrund: Auf der Basis von Schülerangaben wurde eine Aufteilung in zwei Gruppen vorgenommen. Unterschieden wurde zwischen Schülerinnen und Schülern, bei denen kein Elternteil im Ausland geboren wurde (kein Migrationshintergrund), und solchen, bei denen ein Elternteil oder beide im Ausland geboren wurden. Schülerinnen und Schüler ohne Migrationshintergrund wurden für die Analysen mit 0 kodiert, mit Migrationshintergrund erhielten sie den Wert 1.

Geschlecht: Bei der Kodierung des Geschlechts wurde den Mädchen der Wert 1, den Jungen der Wert 0 zugewiesen.

Komposition: Die genannten Hintergrundmerkmale wurden auf der Klassenebene aggregiert, um die soziale, kognitive und ethnische Zusammensetzung der Schulklassen sowie den Mädchenanteil zu kontrollieren.

Schulart: Bei den Schularten wurde zwischen Realschulen, Gymnasien und Schularten mit Hauptschul-Bildungsgang unterschieden (Hauptschulen, Schulen mit mehreren Bildungsgängen und Integrierte Gesamtschulen). Dazu wurden zwei Dummy-Variablen gebildet und die Realschulen als Referenzkategorie herangezogen.

Zielkriterien: Lesekompetenz und Lesefreude

Zielkriterien der folgenden Analysen sind die *Lesefreude* und die *Lesekompetenz* der Schülerinnen und Schüler der neunten Jahrgangsstufe. Die *Lesefreude* der Schülerinnen und Schüler als Indikator für die Lesemotivation wurde über eine Frage zu verschiedenen Aussagen über das eigene Lesen erfasst, die sie auf einer 4-stufigen Antwortskala von „stimmt überhaupt nicht“ bis „stimmt genau“ beantworten konnten. Die Skala mit 11 Items weist eine sehr hohe interne Konsistenz auf ($\alpha = .92$). Ein Beispielitem lautet: „Lesen ist eines meiner liebsten Hobbys.“

Die Lesekompetenz wurde mit dem in PISA 2009 verwendeten Kompetenztest gemessen (OECD, 2011b). Für jede Person wurden fünf *Plausible Values* (PVs; Mislevy, Beaton, Kaplan & Sheehan, 1992) als Personenfähigkeitsparameter generiert. Dies erfolgte unter Verwendung der Itemparameter aus der internationalen PISA-Skalierung, sodass sich die Kompetenzen der Neuntklässlerinnen und Neuntklässler auf die PISA-Metrik beziehen lassen. Für den Bereich Lesen waren in der ersten PISA-Erhebung mit diesem Schwerpunkt, PISA 2000, innerhalb der OECD-Staaten ein Mittelwert von 500 Punkten und eine Standardabweichung von 100 Punkten für die 15-jährigen Schülerinnen und Schüler festgesetzt worden. In der internationalen PISA-2009-Studie lag der OECD-Durchschnitt bei 493 Punkten, die Standardabweichung bei 93 Punkten (Naumann, Artelt, Schneider & Stanat, 2010).

4.3 Methodisches Vorgehen

Die Analysen für diesen Beitrag erfolgten mithilfe von Mehrebenen-Regressionsanalysen (Hox, 2010; Raudenbush & Bryk, 2002; Snijders & Bosker, 2012) unter Nutzung der Software HLM in der Version 6.08 (Raudenbush, Bryk & Congdon, 2009). Unterschieden wurde zwischen drei hierarchisch geschachtelten Ebenen, der Schüler-, der Schulklassen- und der Schulebene. Die für die Auswertung verwendeten Modelle waren *random intercept*-Modelle. Bei diesen kann das Niveau der vorhergesagten Zielvariable systematisch zwischen den Gruppen – hier: Schulklassen und Schulen – variieren. Auf allen drei Ebenen lokalisierte Prädiktoren wurden herangezogen, um Unterschiede zwischen Schülerinnen und Schülern, Klassen und Schulen in den Zielvariablen zu präzisieren. Dabei wurde jede der Zielvariablen in jeweils separaten Analysen untersucht.

Die Schulmerkmale wurden auf einen Mittelwert von 0 und eine Streuung von 1 normiert. Ebenso wurden für die individuellen Schülerwahrnehmungen des Unterrichts, die kontinuierlichen soziobiografischen Hintergrundmerkmale der Schülerinnen und Schüler (HISEI, KFT) sowie für die Lesefreude der Schülerinnen und Schüler ein Mittelwert von 0 und eine Standardabweichung von 1 festgesetzt.

In Mehrebenenanalysen ist die Zentrierung der Prädiktoren von großer Bedeutung (Hofmann & Gavin, 1998; Kreft, de Leeuw & Aiken, 1995). Im vorliegenden Kontext relevant ist die Zentrierung der Hintergrundmerkmale und der Schülerwahrnehmungen des Unterrichts. Die metrischen Hintergrundmerkmale auf Schülerebene (HISEI, KFT) wurden um das Gesamtmittel, die kategorialen Merkmale (Geschlecht, Migrationshin-

tergrund) nicht zentriert. In der Folge lassen sich die Effekte der Hintergrundmerkmale auf Klassenebene als Kompositionseffekte interpretieren (Raudenbush & Bryk, 2002). Die Klassenmittelwerte der Hintergrundmerkmale wurden um das Gesamtmittel zentriert, wodurch die Verschiedenheit der Schulen hinsichtlich der Zusammensetzung ihrer Schülerschaft berücksichtigt wurde (vgl. Raudenbush & Bryk, 2002, S. 142 und S. 250).

Die Schülerwahrnehmungen des Unterrichts wurden auf Schülerebene um das Gruppenmittel (d.h. den Klassenmittelwert) zentriert, auf Klassenebene wurde eine Zentrierung der Mittelwerte um den Gesamtmittelwert vorgenommen. Durch Gruppenzentrierung können Abweichungen der individuellen Schülerwahrnehmungen vom Klassenmittelwert als Nicht-Übereinstimmung mit der geteilten Wahrnehmung der Klasse (dem Mittelwert) aufgefasst werden (vgl. Lüdtke, Robitzsch, Trautwein & Kunter, 2009). Von Interesse waren demnach weniger die kompositionellen als die globalen Beziehungen der Klassenmittelwerte der Schülerwahrnehmungen mit der Lesekompetenz und der Lesefreude. Die Zentrierung der Klassenmittelwerte um das Gesamtmittel bedeutet analog zu den Hintergrundmerkmalen, dass Unterschiede zwischen den Schulen in der Ausprägung der Schülerwahrnehmungen des Unterrichts kontrolliert wurden.

In der sozialwissenschaftlichen Forschung sind fehlende Werte (*missing values*) ein weit verbreitetes Phänomen. Ein für den Umgang mit fehlenden Werten empfohlenes ‚modernes‘ Verfahren ist die Multiple Imputation (Lüdtke, Robitzsch, Trautwein & Köller, 2007; Schafer & Graham, 2002). In der nationalen Erweiterungsstichprobe von PISA 2009 wurde – in Übereinstimmung mit dem Vorgehen bei früheren PISA-Erhebungen – für den sozioökonomischen Status (HISEI) eine Ersetzung fehlender Werte mittels Multipler Imputation vorgenommen. Wegen der Durchführung der Multiplen Imputation sowie der Verwendung von *Plausible Values* (vgl. Abschnitt 4.2, Zielkriterien: Lesekompetenz und Lesefreude) basieren unsere Analysen auf jeweils fünf parallelen Berechnungen, deren Resultate von HLM integriert wurden.

5. Ergebnisse: Bildungsstandardbezogene Kompetenzüberzeugungen und Lerngelegenheiten aus Lehrer- und Schülersicht und ihre Effekte auf Lesekompetenz und Lesefreude

5.1 Individuelle und institutionelle Varianzkomponenten der Lesekompetenz und Lesefreude

Bevor die Unterschiede in der Lesekompetenz und der Lesefreude erklärt werden, wird in einem ersten Analyseschritt geprüft, in welchem Ausmaß die Gesamtvariation der beiden Zielkriterien institutionell (zwischen Schulen und zwischen Schulklassen) und individuell (zwischen Schülerinnen und Schülern) variieren. In Tabelle 1 sind die Varianzanteile der zu erklärenden Variablen auf den drei Ebenen – Individuum, Klasse, Schule – dargestellt. Hinsichtlich der Lernergebnisse zeigen sich die aus vergleichbaren

Studien zu erwartenden Variationsmuster (Baumert & Schümer, 2001; Baumert et al., 2003; Klieme, Steinert & Hochweber, 2010; Prenzel et al., 2004).

Da kumulative schulische Lernergebnisse institutionell durch Selektionseffekte und differenzielle Entwicklungsmilieus zu erklären sind, entfällt bei der Lesekompetenz etwa die Hälfte der Gesamtvarianz auf die institutionelle Ebene Klasse, Schule *und* Schulart, und knapp die Hälfte auf die Individualebene. Wird die Schulart als Indikator für institutionelle Selektionseffekte und differenzielle Entwicklungschancen kontrolliert, reduziert sich die ursprüngliche Varianz der Lesekompetenz um 40%. Die verbleibende Varianz teilt sich dann wie folgt auf: 79% der Varianz in der Lesekompetenz gehen nun auf individuelle Unterschiede zwischen Schülerinnen und Schülern zurück, 6% auf Unterschiede zwischen Klassen und 15% auf Unterschiede zwischen Schulen derselben Schulart. Eine institutionelle Variabilität von insgesamt 21% innerhalb von Schulformen ist durchaus substanziell und indiziert Gestaltungsspielräume bei der Förderung der Lesekompetenz, die durch professionelle Einstellungen und Praktiken in Schule und Unterricht genutzt werden können.

Dagegen entfallen bei der Lesefreude erwartungsgemäß mit 87% der größte Teil der Varianz auf die individuelle Ebene der Schülerinnen und Schüler und die restlichen 13% auf die institutionelle Ebene. Wird die Schulart kontrolliert, reduziert sich die ursprüngliche Gesamtvarianz der Lesefreude lediglich um 8%. Dann gehen 94% der Varianz auf Unterschiede zwischen Schülerinnen und Schülern zurück und nur rund 6% auf institutionelle Unterschiede. Nach diesem Variationsmuster fallen die institutionellen Gestaltungsspielräume für die Förderung der Lesemotivation kleiner aus.

Ebene	Lesekompetenz	Lesefreude
L1: Schüler	47.1	86.6
L2: Klasse	3.4	1.9
L3: Schule und Schulart	49.5	11.6
<i>nach Kontrolle der Schulart</i>		
L1: Schüler	79.0	94.3
L2: Klasse	6.0	2.3
L3: Schule	15.0	3.4

Tab. 1: Individuelle und institutionelle Varianzkomponenten der Lesekompetenz und Lesefreude von Neuntklässlern in PISA 2009 (in %)

5.2 *Kompetenzüberzeugungen und Leseangebote auf Schulebene und Kompetenzüberzeugungen und Nutzung der Leseangebote im Unterricht*

In einem zweiten Analyseschritt wird der Frage nachgegangen, ob und inwieweit die auf die Bildungsstandards bezogenen Kompetenzüberzeugungen im Bereich des Reflektierens und Bewertens und entsprechende Leseangebote der Deutsch-Lehrpersonen auf Schulebene Unterschiede in den Kompetenzüberzeugungen und der Nutzung von Leseangeboten im Unterricht durch die Schülerinnen und Schüler erklären können. Die Ergebnisse der Mehrebenenanalysen sind in Tabelle 2 dargestellt. Für beide Zielvariablen wurden jeweils zwei Modelle gerechnet. In Modell I wurde der Brutto-Effekt der bildungsstandardbezogenen Kompetenzüberzeugungen und Leseangebote der Deutsch-Lehrpersonen auf Schulebene auf die Schülereinschätzungen ermittelt, in Modell II der um die Individualmerkmale der Schülerinnen und Schüler, die Kompositionsmerkmale der Schulklasse und die Schulart korrigierte Netto-Effekt.

Die Ergebnisse in Tabelle 2 zeigen für die Vorhersage der auf das Reflektieren und Bewerten von Texten bezogenen Kompetenzüberzeugungen der Schülerinnen und Schüler und die Vorhersage der Häufigkeit entsprechender Leseaufgaben im Deutschunterricht ein unterschiedliches Befundmuster. Für die Kompetenzüberzeugungen der Schülerinnen und Schüler im Bereich des Reflektierens und Bewertens von Texten zeigen sich im Brutto-Modell I positive Effekte der entsprechenden Kompetenzüberzeugungen und Kompetenzziele der Lehrpersonen des Fachkollegiums Deutsch. Ein Zusammenhang der Häufigkeit der entsprechenden Leseaufgaben im Deutschunterricht sowie der Abstimmung von Zielen und Standards mit den Kompetenzüberzeugungen der Schülerinnen und Schüler zeigt sich nicht. Werden der individuelle Hintergrund, aber vor allem auch die Komposition der Schulklasse und die Schulart kontrolliert, reduziert sich der Effekt der Kompetenzüberzeugungen der Lehrkräfte deutlich, bleibt aber noch bedeutsam (Modell II). Der Effekt der Kompetenzziele verschwindet, wenn die Kontrollvariablen in die Analyse einbezogen werden.

Dass Mädchen, Schülerinnen und Schüler mit höheren kognitiven Fähigkeiten und aus Familien mit hohem sozioökonomischem Status höhere Kompetenzüberzeugungen im Lesen haben, entspricht der allgemeinen Befundlage. Auch in Schulklassen mit privilegierter sozioökonomischer Zusammensetzung zeigen sich – auch bei Kontrolle der Schulart – nicht unerwartet höhere Kompetenzüberzeugungen. Hervorzuheben sind allerdings auch die höheren Kompetenzüberzeugungen in Klassen mit einem höheren Anteil von Schülerinnen und Schülern mit Migrationshintergrund, die – wie weiter unten dargestellt wird – nicht mit höheren Lesekompetenzen einhergehen. Ein Teil der Variation der Kompetenzüberzeugungen der Schülerinnen und Schüler ist erwartungsgemäß auch schulartgebunden, da diese in Gymnasien höher ausgeprägt sind als in den anderen Schularten. Dennoch zeigt sich noch ein – wenn auch recht kleiner – Netto-Effekt der gemeinsam geteilten Kompetenzüberzeugungen der Lehrpersonen einer Schule auf die entsprechenden Kompetenzüberzeugungen in den Klassen.

Bei der Wahrnehmung der Häufigkeit von Leseaufgaben im Bereich des Reflektierens und Bewertens von Texten seitens der Schülerinnen und Schüler zeigen sich im

Prädiktoren	Kompetenzüberzeugung:				Häufigkeit von Leseaufgaben:			
	Reflektieren und Bewerten von Texten				Reflektieren und Bewerten von Texten			
	Modelle	I	(SE)	II	I	(SE)	II	(SE)
Schülebene								
Mädchen		β	(SE)	β	(SE)	β	(SE)	(SE)
mit Migrationshintergrund		0.115	(0.026)			-0.107	(0.026)	
Sozioökonomischer Status: HISEI		-0.027	(0.032)			0.043	(0.031)	
Kognitive Grundfähigkeiten: KFT		0.073	(0.014)			0.022	(0.014)	
		0.091	(0.016)			-0.036	(0.015)	
Klassenebene								
Komposition: Mädchenanteil		0.048	(0.100)			-0.047	(0.133)	
Komposition: mit Migrationshintergrund		0.241	(0.100)			0.359	(0.105)	
Komposition: HISEI		0.168	(0.072)			0.100	(0.065)	
Komposition: KFT		0.042	(0.052)			-0.069	(0.050)	
Schulebene								
Kompetenzüberzeugungen		0.140	(-0.033)			0.019	(0.028)	(0.037)
Kompetenzziele		0.067	(0.032)			-0.008	(0.033)	(0.031)
Leseaufgaben im Deutschunterricht		0.028	(0.027)			-0.012	(0.029)	(0.026)
Absprache von Zielen und Leistungsanforderungen		-0.026	(0.019)			0.043	(0.019)	(0.021)
Schulart: Schulen mit Hauptschulbildungsgang						0.026	(0.058)	
Schulart: Gymnasien						0.065	(0.067)	

Anmerkung: signifikante Regressionskoeffizienten β : fett ($p < .05$); SE: Standardfehler

Tab. 2: Vorhersage der Kompetenzüberzeugungen und wahrgenommenen Leseaufgaben im Deutschunterricht von Schülerinnen und Schülern durch Kompetenzerwartungen und Praktiken der Lehrkräfte des Fachkollegiums Deutsch

Brutto-Modell I überhaupt keine Effekte der entsprechenden Angebote an Leseaufgaben sowie der Kompetenzüberzeugungen und Kompetenzziele der Lehrpersonen. Lediglich die Abstimmung von Zielen und Leistungsanforderungen auf Schulebene steht in schwachem Zusammenhang mit der von den Schülerinnen und Schülern wahrgenommenen Häufigkeit von Leseaufgaben im Deutschunterricht. Dieser Effekt bleibt auch erhalten, wenn die individuellen und kompositionellen Kontrollvariablen sowie die Schulart in die Analysen einbezogen werden (Modell II). Als einziger Kompositionseffekt ist der Migrantenanteil in den Schulklassen für die wahrgenommene Übungsintensität von Leseaufgaben von Bedeutung. Lehrpersonen scheinen demnach in Schulklassen mit einem hohen Migrantenanteil Leseaufgaben im Bereich des Reflektierens und Bewertens häufiger zum Gegenstand des Deutschunterrichts zu machen. Eine hohe Übungsintensität von Leseaufgaben wird auf Individualebene außerdem eher von Jungen und von Schülerinnen und Schülern mit weniger günstigen kognitiven Voraussetzungen wahrgenommen.

5.3 Lesekompetenzen und Lesefreude von Neuntklässlern und Kompetenzüberzeugungen und Leseaufgaben aus Lehrer- und Schülersicht

In einem dritten Analyseschritt wird schließlich geprüft, ob und inwieweit Unterschiede in der Lesekompetenz und Lesefreude durch die Kompetenzüberzeugungen und die Nutzung von Leseaufgaben im Deutschunterricht aus Schülersicht und die entsprechenden Überzeugungen und Praktiken der Lehrpersonen ihrer Schule vorhergesagt werden können. Hierzu wurden jeweils drei Modelle gerechnet: In Modell I wurde der Brutto-Effekt der Kompetenzüberzeugungen und Praktiken des Fachkollegiums Deutsch auf die Lesekompetenz und Lesefreude ermittelt. In Modell II wurde zusätzlich der um die Individualmerkmale der Schülerinnen und Schüler und die Kompositionsmerkmale der Schulklasse korrigierte Netto-Effekt der Kompetenzüberzeugungen und der von den Schülerinnen und Schülern wahrgenommenen Nutzung von Leseaufgaben im Deutschunterricht ermittelt. Diese Effekte wurden in Modell III schließlich noch um die Schularten korrigiert, die als differenzielle Entwicklungs- und Lernmilieus einen nicht unerheblichen Effekt auf Kompetenzerwartungen und Kompetenzen haben. Die Ergebnisse der Analysen sind in Tabelle 3 dargestellt.

Im Modell I zeigen die auf die Bildungsstandards bezogenen Kompetenzüberzeugungen und Kompetenzziele der Lehrkräfte im Bereich des Reflektierens und Bewertens von Texten einen positiven Brutto-Effekt auf die Lesekompetenz im Umfang von etwa 25 Punkten bzw. einem Viertel einer Standardabweichung. In Schulen, die sich verstärkt über Leistungsanforderungen verständigen, weisen Schülerinnen und Schüler je Standardabweichung eine um 12 Punkte niedrigere Lesekompetenz auf. Fehlende Klarheit über Leistungsstandards geht offenbar mit niedrigeren Lesekompetenzen einher.

Prädiktoren	Modelle	Lesekompetenz						Lesefreude					
		I		II		III		I		II		III	
		β	SE	β	SE	β	SE	β	SE	β	SE	β	SE
Schülerebene													
Mädchen				24.570	(1.547)	24.582	(1.548)			0.690	(0.023)	0.690	(0.023)
mit Migrationshintergrund				-10.742	(1.905)	-10.729	(1.904)			0.056	(0.027)	0.056	(0.027)
Sozioökonomischer Status: HISEI				4.026	(0.884)	4.038	(0.885)			0.058	(0.012)	0.058	(0.012)
Kognitive Grundfähigkeiten: KFT				25.603	(0.945)	25.599	(0.944)			0.086	(0.014)	0.086	(0.014)
Kompetenzüberzeugungen				13.735	(0.877)	13.735	(0.877)			0.220	(0.013)	0.220	(0.013)
Leseaufgaben im Deutschunterricht				-5.343	(0.840)	-5.342	(0.840)			-0.002	(0.012)	-0.002	(0.012)
Klassenebene													
Komposition: Mädchen				-0.810	(8.833)	0.809	(8.199)			0.071	(0.089)	0.077	(0.090)
Komposition:				-3.672	(8.817)	-6.536	(9.329)			0.227	(0.072)	0.228	(0.071)
mit Migrationshintergrund													
Komposition: HISEI				22.736	(4.228)	13.904	(4.187)			0.311	(0.054)	0.307	(0.057)
Komposition: KFT				28.898	(3.505)	21.923	(3.681)			0.050	(0.042)	0.053	(0.045)
Kompetenzüberzeugungen				21.274	(4.958)	19.100	(4.868)			0.208	(0.052)	0.206	(0.053)
Leseaufgaben im Deutschunterricht				-9.950	(3.880)	-9.327	(3.671)			0.004	(0.040)	0.003	(0.040)
Schulebene (Fachkollegium Deutsch)													
Kompetenzüberzeugungen		23.889	(6.977)	4.737	(1.968)	3.483	(1.830)	0.142	(0.033)	0.015	(0.016)	0.016	(0.017)
Kompetenzziele		25.216	(7.315)	5.195	(2.462)	3.385	(2.812)	0.114	(0.038)	0.005	(0.021)	0.008	(0.021)
Leseaufgaben im Deutschunterricht		-0.252	(5.264)	-2.043	(1.814)	-2.395	(1.884)	0.029	(0.030)	0.017	(0.019)	0.015	(0.019)
Absprache von Zielen und Leistungsanforderungen		-11.666	(3.248)	-2.997	(1.470)	-2.043	(1.445)	-0.084	(0.020)	-0.034	(0.014)	-0.034	(0.015)
Schulart: Schulen mit Hauptschulbildungsgang						-10.682	(5.602)			0.025		0.025	(0.044)
Schulart: Gymnasien						19.855	(3.886)			0.022		0.022	(0.045)

Anmerkung: signifikante Regressionskoeffizienten β : fett ($p < .05$); SE: Standardfehler

Tab. 3: Vorhersage der Lesekompetenz und Lesefreude durch die Kompetenzüberzeugungen und Lesesaufgaben im Deutschunterricht aus Lehrer- und Schülersicht

Werden in Modell II die Kompetenzüberzeugungen und die Nutzung von Leseaufgaben im Deutschunterricht aus Sicht der Schülerinnen und Schüler sowie ihre individuellen und kompositionellen Hintergrundmerkmale in die Analyse mit einbezogen, reduzieren sich die Effekte der Kompetenzüberzeugungen und Kompetenzziele der Lehrkräfte auf die Lesekompetenz auf 5 Punkte, der Absprache im Kollegium auf 3 Punkte. Die Zusammenhänge mit der Lesekompetenz sind aber immer noch signifikant. Die gemeinsam geteilten Kompetenzüberzeugungen der Schülerinnen und Schüler in den Klassen zeigen einen positiven Effekt auf die Lesekompetenz im Umfang von 21 Punkten bzw. einem Fünftel einer Standardabweichung. Die Übungsintensität von Leseaufgaben weist einen negativen Effekt auf die Lesekompetenz im Umfang von 10 Punkten auf. Auch die Komposition der Schulklassen nach sozioökonomischem Status und kognitiven Grundvoraussetzungen ist mit 23 bzw. 29 Punkten für die Vorhersage von Unterschieden in der Lesekompetenz bedeutsam.

Wird schließlich in Modell III die Schularart als Kontrollvariable für differenzielle Lernmilieus eingeführt, sind die Effekte der Lehrerüberzeugungen und Praktiken auf Schulebene nicht mehr bedeutsam, sondern nur die Schularart. Die Lesekompetenzen in Gymnasien liegen um 20 Punkte höher als in Realschulen (Referenzkategorie). Mit der Einführung der Schularart in das Modell reduzieren sich außerdem die sozialen und kognitiven Kompositionseffekte deutlich auf (signifikante) 14 bzw. 22 Punkte.

Die Effekte der gemeinsam geteilten Kompetenzüberzeugungen der Schülerinnen und Schüler und der von ihnen wahrgenommenen Übungsintensität von Leseaufgaben im Umfang von 19 bzw. 9 Punkten verändern sich auch bei Kontrolle der Klassenkomposition und Schularart nur minimal. Der negative Zusammenhang zwischen Übungsintensität und Lesekompetenz lässt bei Kontrolle von Kompositionsmerkmalen und Schularart wie vermutet auf ein remediales Unterrichtsverhalten der Lehrkräfte schließen.

Die Effekte der individuellen Hintergrundmerkmale auf die Lesekompetenz zeigen alle die erwartete Richtung: Mädchen und Schülerinnen und Schüler mit günstigeren kognitiven Grundfähigkeiten sowie mit höherem sozioökonomischen Status weisen deutlich bzw. geringfügig höhere Lesekompetenzen auf, Schülerinnen und Schüler mit Migrationshintergrund zeigen niedrigere Lesekompetenzen. Die Effekte der individuellen Kompetenzüberzeugungen und Unterrichtswahrnehmungen von Leseaufgaben seitens der Schülerinnen und Schüler zeigen das gleiche Muster wie die Einschätzungen auf Klassenebene.

In Tabelle 3 ebenfalls dargestellt sind die Ergebnisse der Mehrebenenanalysen zur Vorhersage der Lesefreude. Für die Analysen werden analog zur Vorgehensweise bei der Lesekompetenz drei Modelle gerechnet.

Im Modell I weisen die Kompetenzüberzeugungen und Kompetenzziele der Lehrkräfte einen positiven Brutto-Effekt auf die Lesefreude im Umfang von 14% und 11% einer Standardabweichung auf. In Schulen, die sich verstärkt über Leistungsanforderungen verständigen, resultiert dagegen eine etwas geringere Lesefreude.

Werden in Modell II die Kompetenzüberzeugungen und die Häufigkeit von Leseaufgaben im Deutschunterricht aus Sicht der Schülerinnen und Schüler sowie die individuellen und kompositionellen Hintergrundmerkmale und in Modell III zusätzlich noch

die Schulart in die Analyse einbezogen, sind die Effekte der Kompetenzüberzeugungen und -ziele der Lehrkräfte auf Schulebene auf die Lesefreude nicht mehr signifikant. Nur die Absprache von Standards und Leistungsanforderungen zeigt noch einen minimalen, aber signifikanten negativen Effekt auf die Lesefreude. Anders als bei der Lesekompetenz weist die Schulart keine Relevanz für Unterschiede in der Lesefreude der Schülerinnen und Schüler auf. Die soziale Komposition der Schulklassen hat auch bei Kontrolle der Schulart einen deutlich positiven Effekt auf die Lesefreude, ebenso der Anteil der Schülerinnen und Schüler mit Migrationshintergrund.

Auf Klassenebene sind der positive Effekt der Kompetenzüberzeugungen und der fehlende Effekt der Übungsintensität bei den Leseaufgaben auf die Lesefreude hervorzuheben. Erwartungsgemäß sind die von den Schülerinnen und Schülern einer Klasse geteilten Kompetenzüberzeugungen nicht nur mit der Lesekompetenz, sondern auch mit der Lesefreude positiv verknüpft. Eine höhere Übungsintensität scheint die Einstellung zum Lesen nicht zu beeinträchtigen und nicht zu befördern. Die Effekte der individuellen Hintergrundmerkmale auf die Lesefreude zeigen alle die erwartete Richtung: Mädchen, Schülerinnen und Schüler mit höheren kognitiven Grundfähigkeiten sowie einem höheren sozioökonomischen Status, und schließlich Schülerinnen und Schüler mit Migrationshintergrund weisen eine höhere Lesefreude auf. Die individuellen Kompetenzüberzeugungen stehen ebenfalls in positiver Beziehung zur Lesefreude. Die wahrgenommene Übungsintensität von Leseaufgaben ist auch auf Individualebene für die Lesefreude nicht bedeutsam.

6. Zusammenfassung und Diskussion

Ziel dieses Beitrags war es, anhand der national erweiterten Stichprobe von PISA 2009 zu klären, ob und inwieweit bildungsstandardbezogene Kompetenzüberzeugungen von Lehrkräften und Schülerinnen und Schülern im Bereich des Lesens sowie das Angebot und die Nutzung von Leseaufgaben im Deutschunterricht auf Schul- bzw. Unterrichtsebene Unterschiede in der Lesekompetenz und der Lesefreude der Jugendlichen vorhersagen können. Im Folgenden werden die Befunde aus den Mehrebenenanalysen im Hinblick auf die Relevanz der Ergebnisse für die Schul- und Unterrichtseffektivitätsforschung und die Schulpraxis diskutiert. (1) Wie groß sind die individuelle und institutionelle Variabilität der Lesekompetenz und Lesefreude, die als Ausgangsbasis für die Gestaltung von Schule und Unterricht genutzt werden können? (2) Gibt es Zusammenhänge zwischen den Schul- und Unterrichtsmerkmalen, zwischen den gemeinsam geteilten Überzeugungen, Angeboten und Praktiken der Lehrpersonen in Schulen und den entsprechenden Überzeugungen und der Nutzung von Angeboten durch die Schülerinnen und Schüler im Unterricht? (3) Gibt es einen Zusammenhang zwischen den Schul- und Unterrichtsmerkmalen und den Lesekompetenzen und der Lesefreude der Schülerinnen und Schüler? (4) Welchen Grenzen unterliegt die Untersuchung und welche Fragen bleiben offen?

(1) In einem ersten Analyseschritt wurde das Ausmaß der individuellen und institutionellen Variabilität der beiden Zielkriterien, Lesekompetenz und Lesefreude, ermittelt. Bei der Lesekompetenz der Schülerinnen und Schüler zeigten sich in etwa die aus früheren Studien bekannten Befunde: Solange die Schulart als Indikator für differenzielle Lernmilieus in einem gegliederten Schulsystem nicht auspartialisiert wird, entfällt knapp die Hälfte der Gesamtvarianz der Schülerkompetenzen im Lesen auf die Schulebene. Wird die Schulart kontrolliert, gehen immer noch 15% der verbleibenden Varianz auf Unterschiede zwischen Schulen und 6% auf Unterschiede zwischen Klassen zurück. Bei der Lernfreude entfallen 13% der Gesamtvarianz auf institutionelle Unterschiede zwischen Schulen und Schulklassen und 87% auf individuelle Unterschiede zwischen Schülerinnen und Schülern. Unter Kontrolle der Schulart reduziert sich der Anteil der institutionellen Varianz an der Gesamtvarianz auf 6%. Dieses Ausmaß institutioneller Variabilität weist darauf hin, dass die Gestaltungsspielräume von Schule und Unterricht für die Förderung der Lesekompetenz größer sind als für die Lesefreude.

(2) Die Frage, ob und inwieweit die auf die Bildungsstandards im Lesen bezogenen Kompetenzüberzeugungen und Praktiken im Fachkollegium Deutsch auf Schulebene Unterschiede in den Kompetenzüberzeugungen und der wahrgenommenen Übung von Leseaufgaben im Deutschunterricht aus Schülersicht erklären können, wurde in einem zweiten Schritt anhand von Mehrebenenanalysen geprüft. Im Brutto-Modell mit den Schulmerkmalen als einzigen Prädiktoren sind die gemeinsam geteilten Kompetenzüberzeugungen und Absprachen von Leistungsanforderungen der Lehrkräfte, nicht aber das Angebot von Leseaufgaben für Unterschiede in den entsprechenden Kompetenzüberzeugungen der Schülerinnen und Schüler von Bedeutung. Dieser Brutto-Effekt der Lehrerüberzeugungen auf die Schülerüberzeugungen wird deutlich reduziert, wenn die Schulart und die individuellen und kompositionellen Hintergrundmerkmale der Schülerinnen und Schüler in die Analyse einbezogen werden. Ein schwacher Zusammenhang zwischen den Kompetenzüberzeugungen der Lehrpersonen und der Schülerinnen und Schüler bleibt allerdings bestehen. Dieser Zusammenhang deutet darauf hin, dass Kompetenzüberzeugungen auf der Ebene von Schulen Aspekte einer Schulkultur sind, welche die Erwartungshorizonte auch von Schülerinnen und Schülern im Unterricht erweitern könnten. Dass sich die Entwicklung einer anspruchsvollen und differenzierten Aufgabenkultur im Unterricht für das professionelle Lernen von Lehrpersonen und für die Verbesserung von Schülerkompetenzen lohnen kann, hat das Modellprogramm SINUS für die Mathematik gezeigt (Prenzel, Carstensen, Senkbeil, Ostermeier & Seidel, 2005; Prenzel & Fischer, 2009).

Die Bedeutung der Schulart für die bildungsstandardbezogenen Kompetenzüberzeugungen von Lehrkräften hängt möglicherweise auch damit zusammen, dass die Bildungsstandards der KMK in den verschiedenen Fächern nicht nur mit Bezug auf fachspezifische Anforderungen und Kompetenzmodelle, sondern auch abschlussbezogen bzw. bildungsgangspezifisch formuliert worden sind. Schulartspezifische Ausprägungen von Kompetenzüberzeugungen und Praktiken könnten auch mit Unterschieden in der Lehramtsausbildung zusammenhängen, die nicht nur Unterschiede im Fachwissen und fachdidaktischen Wissen der Lehrkräfte (Brunner et al., 2006), sondern auch in den

kognitiven Zielsetzungen und Bewertungsmaßstäben zur Folge haben (vgl. Hochweber, 2010; Kunter et al., 2005).

Die fehlenden Zusammenhänge zwischen der Übungsintensität von Leseaufgaben auf Schul- und Unterrichtsebene könnten ein Hinweis auf ein differenziertes und adaptives Verhalten von Lehrkräften im Hinblick auf unterschiedliche Schülervoraussetzungen in den Klassen sein. Dafür spricht auch eine hohe wahrgenommene Übungsintensität von Leseaufgaben in Klassen mit hohem Anteil von Schülerinnen und Schülern mit Migrationshintergrund. Zur Prüfung der spezifischen Zusammenhänge der Kompetenzüberzeugungen und der Lerngelegenheiten auf Schul- und Unterrichtsebene wären jedoch bildungsgang- bzw. schulartspezifische Analysen erforderlich. Dafür stehen allerdings nicht genügend Schulen innerhalb der Schularten in der Stichprobe zur Verfügung.

(3) Die Vorhersage der Lesekompetenz und Lesefreude durch die geteilten Kompetenzüberzeugungen und Praktiken der Lehrkräfte sowie die geteilten Kompetenzüberzeugungen und die wahrgenommene Übungsintensität aus Sicht der Schülerinnen und Schüler bestätigt die in Abschnitt 2 und 3 postulierten Zusammenhänge weitgehend. Da sich der im Brutto-Modell ermittelte Zusammenhang zwischen den Kompetenzüberzeugungen und Kompetenzzielen der Lehrkräfte mit der Lesekompetenz und Lesefreude bei Kontrolle individueller und kompositioneller Hintergrundmerkmale deutlich verringert und nach Einbeziehung der Schulart ganz verschwindet, gibt es keine direkten Effekte der Schulmerkmale auf die Lernergebnisse der Schülerinnen und Schüler. Auch die Leseangebote auf Schulebene sind für Unterschiede in der Lesekompetenz und Lesefreude nicht bedeutsam. Nur der schwache Zusammenhang zwischen einer erhöhten Absprache von Leistungsanforderungen und der Lesefreude der Schülerinnen und Schüler deutet darauf hin, dass verstärkte Bemühungen um eine Klarheit in den Leistungsanforderungen in der Schule auch bei Berücksichtigung der Klassenzusammensetzung und Schulart mit ungünstigen Motivationslagen zusammenfallen.

Die gemeinsam geteilten Kompetenzüberzeugungen der Schülerinnen und Schüler auf der Ebene des Unterrichts sind – auch bei Kontrolle der kognitiven und sozialen Klassenkomposition und der Schulart – für Unterschiede in der Lesekompetenz und Lesefreude bedeutsam. Dieser Befund entspricht nicht nur den theoretisch angenommenen Zusammenhängen, er ist auch praktisch bedeutsam. Da bei Schülerinnen und Schülern mit hohen Kompetenzüberzeugungen im Lesen davon auszugehen ist, dass sie in ihren Klassen positive Erfahrungen im Lesen und in der Bewältigung anspruchsvoller Texte gemacht haben und über die für die Bewältigung der Leseanforderungen erforderlichen Mittel auch eher verfügen, ist dieses Potential auch im Unterricht nutzbar. Kompetenzüberzeugungen können Lehrkräfte im Unterricht gut fördern, indem Schülerinnen und Schüler z.B. ihre Verantwortlichkeit für Leseerfolge und -fortschritte erleben, effektive Lesestrategien nutzen, Erfahrungen mit vielfältigen und unterschiedlich schwierigen Texten sammeln und konstruktive, aufgabenspezifische Rückmeldungen erhalten (Köller & Möller, 2010). Der positive Zusammenhang zwischen den geteilten Kompetenzüberzeugungen der Lehrkräfte und den geteilten Kompetenzüberzeugungen der Schülerinnen und Schüler deutet zudem darauf hin, dass Aspekte der Schul-

kultur für die Förderung von Kompetenzüberzeugungen im Unterricht bedeutsam sein könnten.

Da zwischen der Übungsintensität der auf die Bildungsstandards bezogenen Leseaufgaben im Deutschunterricht und den Lesekompetenzen ein negativer Zusammenhang besteht, wird ein remediales Unterrichtsverhalten der Lehrkräfte angenommen. Die ermittelten Effekte der kognitiven und sozialen Zusammensetzung auf die Lesekompetenz und der sozialen Komposition auf die Lesefreude sind nicht überraschend. Hervorzuheben ist die höhere Lesefreude in Klassen mit einem höheren Anteil von Schülerinnen und Schülern mit Migrationshintergrund. In diesen Klassen dürften zwar auch Bezugsgruppeneffekte für die Wahrnehmung und Bewertung der eigenen Erfolgsaussichten und die Motivation von Bedeutung sein (Köller, Schnabel & Baumert, 2000), aber als lernförderlicher Klassenkontext wäre dieser Zusammenhang gleichwohl nutzbar.

(4) Die in diesem Beitrag dargestellten Befunde und Interpretationen sind auch vor dem Hintergrund der Grenzen dieser Untersuchung zu werten, weshalb auch einige Interpretationen letztlich als offene Fragen formuliert werden müssen. Das Querschnittsdesign von PISA mit nur einem Messzeitpunkt ermöglicht es nicht, die Richtung von Effekten, also auch die Annahmen über remediale Beziehungen, zu überprüfen. Damit können auch die Veränderungen in den Lernergebnissen der Schülerinnen und Schüler durch sich verändernde – adaptive – Schul- und Unterrichtsmerkmale nicht erfasst werden. Die Frage nach der zeitlichen Stabilität der Schul- und Unterrichtsmerkmale und ihrer Beziehung zu den Schülerkompetenzen muss danach offen bleiben. Außerdem kann angesichts der geringen Anzahl von Schulen je Schulart nicht geprüft werden, inwieweit die ermittelten Zusammenhänge zwischen den Schul- und Unterrichtsmerkmalen und zwischen den Unterrichtsmerkmalen und den Lernergebnisse der Schülerinnen und Schüler in den einzelnen Schularten gleich sind. Schließlich war die Untersuchung der Effekte der Kompetenzüberzeugungen und Praktiken in Schule und Unterricht auf das Lesen im Fach Deutsch beschränkt, auch wenn davon ausgegangen wird, dass auch andere Fächer und das außerschulische Lernen gerade für die Lesekompetenz von Bedeutung sind. Gleichwohl können die gefundenen Zusammenhänge zumindest als Ausgangspunkt für die lesebezogene Entwicklung des Unterrichts und das professionelle Lernen von Lehrkräften in den Schulen genutzt werden.

Literatur

- Artelt, C., McElvany, N., Christmann, U., Richter, T., Groeben, N., Köster, J., Schneider, W., Stanat, P., Ostermeier, C., Schiefele, U., Valtin, R., & Ring, K. (2005). *Förderung von Lesekompetenz. Eine Expertise*. Bonn: Bundesministerium für Bildung und Forschung (BMBF).
- Artelt, C., Naumann, J., & Schneider, W. (2010). Lesemotivation und Lernstrategien. In E. Klieme, C. Artelt, J. Hartig, N. Jude, O. Köller, M. Prenzel, W. Schneider & P. Stanat (Hrsg.), *PISA 2009. Bilanz nach einem Jahrzehnt* (S. 73-112). Münster: Waxmann.
- Bandura, A. (1977). *Social learning theory*. Englewood Cliffs: Prentice-Hall.
- Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York: Freeman.

- Baumert, J., Kunter, M., Blum, W., Brunner, M., Voss, T., Jordan, A., Klusmann, U., Krauss, S., Neubrand, M., & Tsai, Y.-M. (2010). Teachers' mathematical knowledge, cognitive activation in the classroom, and student progress. *American Educational Research Journal*, 47, 133-180.
- Baumert, J., & Schümer, G. (2001). Schulformen als selektionsbedingte Lernmilieus. In Deutsches PISA Konsortium (Hrsg.), *PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich* (S. 454-467). Opladen: Leske + Budrich.
- Baumert, J., & Schümer, G. (2002). Familiäre Lebensverhältnisse, Bildungsbeteiligung und Kompetenzerwerb im nationalen Vergleich. In Deutsches PISA Konsortium (Hrsg.), *PISA 2000. Die Länder der Bundesrepublik Deutschland im Vergleich* (S. 159-202). Opladen: Leske + Budrich.
- Baumert, J., Trautwein, U., & Artelt, C. (2003). Schulumwelten – institutionelle Bedingungen des Lehrens und Lernens. In J. Baumert, C. Artelt, E. Klieme, M. Neubrand, M. Prenzel, U. Schiefele, W. Schneider, K.-J. Tillmann & M. Weiß (Hrsg.), *PISA 2000. Ein differenzierter Blick auf die Länder der Bundesrepublik Deutschland* (S. 261-331). Opladen: Leske + Budrich.
- Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin*, 88, 588-606.
- Bloom, B. S. (1976). *Human characteristics and school learning*. New York: McGraw-Hill.
- Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York: Guilford.
- Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Hrsg.), *Testing structural equation models* (S. 136-162). Newbury Park: Sage.
- Brunner, M., Kunter, M., Krauss, S., Baumert, J., Blum, W., Dubberke, T., Jordan, A., Klusmann, U., Tsai, Y.-M., & Neubrand, M. (2006). Welche Zusammenhänge bestehen zwischen dem fachspezifischen Professionswissen von Mathematiklehrkräften und ihrer Ausbildung sowie beruflichen Fortbildung? *Zeitschrift für Erziehungswissenschaft*, 9(4), 521-544.
- Carroll, J. (1963). A model for school learning. *Teachers College Record*, 64, 723-733.
- Chrispeels, J. H., & Harris, A. (2006). Conclusion: Future directions for the field. In A. Harris & J. H. Chrispeels (Hrsg.), *Improving schools and educational systems* (S. 295-307). London: Routledge.
- Creemers, B. P. M., & Kyriakides, L. (2008). *The dynamics of educational effectiveness: A contribution to policy, practice and theory in contemporary schools*. London: Routledge.
- Ditton, H. (2000). Qualitätskontrolle und -sicherung in Schule und Unterricht – ein Überblick zum Stand der empirischen Forschung. *Zeitschrift für Pädagogik*, 41. Beiheft, 73-92.
- Ditton, H. (2007). Schulqualität. Modelle zwischen Konstruktion, empirischen Befunden und Implementierung. In J. van Buer & C. Wagner (Hrsg.), *Qualität von Schule. Ein kritisches Handbuch* (S. 83-92). Frankfurt a.M.: Peter Lang.
- Ditton, H., & Krecker, L. (1995). Qualität von Schule und Unterricht. Empirische Befunde zu Fragestellungen und Aufgaben der Forschung. *Zeitschrift für Pädagogik*, 41(4), 507-531.
- Dreeben, R., & Barr, R. (1988). Classroom composition and the design of instruction. *Sociology of Education*, 61(3), 129-142.
- Edmonds, R. (1979). Effective schools for the urban poor. *Educational Leadership*, 37(1), 15-27.
- Ehmke, T., & Baumert, J. (2007). Soziale Herkunft und Kompetenzerwerb. Vergleiche zwischen PISA 2000, 2003 und 2006. In M. Prenzel, C. Artelt, J. Baumert, W. Blum, M. Hammann, E. Klieme & R. Pekrun (Hrsg.), *PISA 2006. Die Ergebnisse der dritten internationalen Vergleichsstudie* (S. 309-335). Münster: Waxmann.
- Ehmke, T., Hohensee, F., Heidemeier, H., & Prenzel, M. (2004). Familiäre Lebensverhältnisse, Bildungsbeteiligung und Kompetenzerwerb. In M. Prenzel, J. Baumert, W. Blum, R. Lehmann, D. Leutner, M. Neubrand, R. Pekrun, H.-G. Rolff, J. Rost & U. Schiefele (Hrsg.), *PISA 2003. Der Bildungsstand der Jugendlichen in Deutschland – Ergebnisse des zweiten internationalen Vergleichs* (S. 225-282). Münster: Waxmann.

- Ehmke, T., & Jude, N. (2010). Soziale Herkunft und Kompetenzerwerb. In E. Klieme, C. Artelt, J. Hartig, N. Jude, O. Köller, M. Prenzel, W. Schneider & P. Stanat (Hrsg.), *PISA 2009. Bilanz nach einem Jahrzehnt* (S. 277-300). Münster: Waxmann.
- Fend, H. (2006). *Neue Theorie der Schule*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Fend, H. (2008). *Schule gestalten. Systemsteuerung, Schulentwicklung und Unterrichtsqualität*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Fraser, B. J., Walberg, H. J., Welch, W. W., & Hattie, J. A. (1987). Syntheses of educational productivity research. *International Journal of Educational Research*, 11, 145-252.
- Ganzeboom, H. B. G., de Graaf, P. M., Treiman, D. J., & de Leeuw, J. (1992). A standard international socio-economic index of occupational status. *Social Science Research*, 21, 1-56.
- Gruehn, S. (2000). *Unterricht und schulisches Lernen. Schüler als Quellen der Unterrichtsbeschreibung*. Münster: Waxmann.
- Hattie, J. A. (2009). *Visible learning. A synthesis of over 800 meta-analyses relating to achievement*. London: Routledge.
- Heller, K. A., & Perleth, C. (2000). *Kognitiver Fähigkeitstest für 4. bis 12. Klassen. Revision*. Göttingen: Beltz.
- Helmke, A., & Weinert, F. E. (1997). Unterrichtsqualität und Leistungsentwicklung: Ergebnisse aus dem SCHOLASTIK-Projekt. In F. E. Weinert & A. Helmke (Hrsg.), *Entwicklung im Grundschulalter* (S. 241-251). Weinheim: Beltz/PVU.
- Hertel, S., Hochweber, J., Steinert, B., & Klieme, E. (2010). Schulische Rahmenbedingungen und Lerngelegenheiten im Deutschunterricht. In E. Klieme, C. Artelt, J. Hartig, N. Jude, O. Köller, M. Prenzel, W. Schneider & P. Stanat (Hrsg.), *PISA 2009. Bilanz nach einem Jahrzehnt* (S. 113-151). Münster: Waxmann.
- Hertel, S., Steinert, B., Mildner, D., Rauch, D., et al. (in Vorb.). *PISA 2009: Dokumentation der Erhebungsinstrumente*.
- Hochweber, J. (2010). *Was erfassen Mathematiknoten? Korrelate von Mathematik-Zeugnissen auf Schüler- und Schulklassenebene in Primar- und Sekundarstufe*. Münster: Waxmann.
- Hochweber, J., Steinert, B., & Klieme, E. (2012). Lehrerkoooperation, Unterrichtsqualität und Lernergebnisse im Fach Englisch. *Unterrichtswissenschaft*, 40(4), 351-370.
- Hofmann, D. A., & Gavin, M. B. (1998). Centering decisions in hierarchical linear models: Implications for research in organizations. *Journal of Management*, 24(5), 623-641.
- Hox, J. J. (2010). *Multilevel analysis: Techniques and applications*. New York: Routledge.
- Hu, L.-T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6(1), 1-55.
- Klieme, E., Artelt, C., Hartig, J., Jude, N., Köller, O., Prenzel, M., Schneider, W., & Stanat, P. (Hrsg.) (2010). *PISA 2009. Bilanz nach einem Jahrzehnt*. Münster: Waxmann.
- Klieme, E., Jude, N., Baumert, J., & Prenzel, M. (2010). PISA 2000-2009: Bilanz der Veränderungen im Schulsystem. In E. Klieme, C. Artelt, J. Hartig, N. Jude, O. Köller, M. Prenzel, W. Schneider & P. Stanat (Hrsg.), *PISA 2009. Bilanz nach einem Jahrzehnt* (S. 277-300). Münster: Waxmann.
- Klieme, E., Jude, N., Rauch, D., Ehlers, H., Helmke, A., Eichler, W., Thomé, G., & Willenberg, H. (2008). Alltagspraxis, Qualität und Wirksamkeit des Deutschunterrichts. In E. Klieme, A. Helmke, R. Lehmann, G. Nold, H.-G. Rolff, K. Schröder et al. (Hrsg.), *Unterricht und Kompetenzerwerb in Deutsch und Englisch. Ergebnisse der DESI-Studie* (S. 313-344). Weinheim/Basel: Beltz.
- Klieme, E., Pauli, C., & Reusser, K. (2009). The Pythagoras Study. In T. Janik & T. Seidel (Hrsg.), *The power of video studies in investigating teaching and learning in the classroom* (S. 137-160). Münster: Waxmann.
- Klieme, E., & Rakoczy, K. (2003). Unterrichtsqualität aus Schülerperspektive: Kulturspezifische Profile, regionale Unterschiede und Zusammenhänge mit Effekten von Unterricht. In J. Baumert, C. Artelt, E. Klieme, M. Neubrand, M. Prenzel, U. Schiefele, W. Schneider &

- K.-J. Tillmann (Hrsg.), *PISA 2000. Ein differenzierter Blick auf die Länder der Bundesrepublik Deutschland* (S. 334-359). Opladen: Leske + Budrich.
- Klieme, E., Schümer, G., & Knoll, S. (2001). Mathematikunterricht in der Sekundarstufe I: Aufgabenkultur und Unterrichtsgestaltung. In Bundesministerium für Bildung und Forschung (BMBF) (Hrsg.), *TIMSS – Impulse für Schule und Unterricht. Forschungsbefunde, Reform-initiativen, Praxisberichte und Video-Dokumente* (S. 43-57). München: Medienhaus Bering.
- Klieme, E., Steinert, B., & Hochweber, J. (2010). Zur Bedeutung der Schulqualität für Unterricht und Lernergebnisse. In W. Bos, E. Klieme & O. Köller (Hrsg.), *Schulische Lerngelegenheiten und Kompetenzentwicklung. Festschrift für Jürgen Baumert* (S. 231-255). Münster: Waxmann.
- Köller, O., & Möller, J. (2010). Selbstwirksamkeit. In D. H. Rost (Hrsg.), *Handwörterbuch Pädagogische Psychologie* (4. Aufl., S. 767-774). Weinheim: Beltz/PVU.
- Köller, O., Schnabel, K., & Baumert, J. (2000). Der Einfluss der Leistungsstärke von Schulen auf das fachspezifische Selbstkonzept der Begabung und das Interesse. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 32, 70-80.
- KMK (2003). *Vereinbarung über Bildungsstandards für den Mittleren Schulabschluss (Jahrgangsstufe 10)*.
http://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2003/2003_12_04-Bildungsstandards-Mittleren-SA.pdf [23.05.2012].
- KMK (2004). *Vereinbarung über Bildungsstandards für den Hauptschulabschluss (Jahrgangsstufe 9)*.
http://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2004/2004_10_15-Bildungsstandards-Haupt.pdf [23.05.2012].
- Kreft, I. G. G., de Leeuw, J., & Aiken, L. S. (1995). The effect of different forms of centering in hierarchical linear models. *Multivariate Behavioral Research*, 30(1), 1-21.
- Kunter, M., Brunner, M., Baumert, J., Klusmann, U., Krauss, S., Blum, W., Jordan, A., & Neubrand, M. (2005). Der Mathematikunterricht der PISA-Schülerinnen und -Schüler. Schulformunterschiede in der Unterrichtsqualität. *Zeitschrift für Erziehungswissenschaft*, 8(4), 502-520.
- Lüdtke, O., Robitzsch, A., Trautwein, U., & Köller, O. (2007). Umgang mit fehlenden Werten in der psychologischen Forschung: Probleme und Lösungen. *Psychologische Rundschau*, 58(2), 103-117.
- Lüdtke, O., Robitzsch, A., Trautwein, U., & Kunter, M. (2009). Assessing the impact of learning environments: How to use student ratings of classroom and school characteristics in multilevel modeling. *Contemporary Educational Psychology*, 34, 120-131.
- McElvany, N., & Becker, M. (2010). Welche Prädiktoren braucht man zur Vorhersage von Lesekompetenz? Eine Kommunalitätsanalyse zur Bestimmung der unigen und geteilten Varianzaufklärung psychologischer und soziologischer Konstrukte. In W. Bos, E. Klieme & O. Köller (Hrsg.), *Schulische Lerngelegenheiten und Kompetenzentwicklung. Festschrift für Jürgen Baumert* (S. 147-172). Münster: Waxmann.
- Mislevy, R. J., Beaton, A. E., Kaplan, B., & Sheehan, K. M. (1992). Estimating population characteristics from sparse matrix samples of items. *Journal of Educational Measurement*, 29(2), 133-164.
- Naumann, J., Artelt, C., Schneider, W., & Stanat, P. (2010). Lesekompetenz von PISA 2000 bis PISA 2009. In E. Klieme, C. Artelt, J. Hartig, N. Jude, O. Köller, M. Prenzel, W. Schneider & P. Stanat (Hrsg.), *PISA 2009. Bilanz nach einem Jahrzehnt* (S. 23-71). Münster: Waxmann.
- OECD (2001). Familiärer Hintergrund und Schülerleistungen. In *Lernen für das Leben. Erste Ergebnisse von PISA 2000*. Paris: OECD.
- OECD (2010). *Overcoming social background. Equity in learning opportunities and outcomes*. Paris: OECD.

- OECD (2011a). *PISA 2009 Ergebnisse: Was macht eine Schule erfolgreich? – Lernumfeld und schulische Organisation in PISA (Band IV)*. Paris: OECD.
- OECD (2011b). *PISA 2009. Technical Report*. Paris: OECD.
- Prenzel, M., Carstensen, C. H., Senkbeil, M., Ostermeier, C., & Seidel, T. (2005). Wie schneiden SINUS-Schulen bei PISA ab? Ergebnisse der Evaluation eines Modellversuchsprogramms. *Zeitschrift für Erziehungswissenschaft*, 8(4), 487-501.
- Prenzel, M., & Fischer, C. (2009). Lehrkräfte lernen in Gruppen und Organisationen – Erfahrungen aus zehn Jahren SINUS-Modellversuchen. In O. Zlatkin-Troitschanskaia, K. Beck, D. Sembill, R. Nickolaus & R. Mulder (Hrsg.), *Lehrprofessionalität – Bedingungen, Genese, Wirkungen und ihre Messung* (S. 577-588). Weinheim/Basel: Beltz.
- Prenzel, M., Senkbeil, M., & Drechsel, B. (2004). Kompetenzunterschiede zwischen Schulen. In Deutsches PISA Konsortium (Hrsg.), *PISA 2003. Der Bildungsstand der Jugendlichen in Deutschland – Ergebnisse des zweiten internationalen Vergleichs* (S. 292-296). Münster: Waxmann.
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods*. Newbury Park: Sage.
- Raudenbush, S. W., Bryk, A. S., & Congdon, R. (2009). *HLM 6. Hierarchical linear and non-linear modeling*. Lincolnwood: SSI.
- Rolff, H.-G. (2007). *Studien zu einer Theorie der Schulentwicklung*. Weinheim/Basel: Beltz.
- Rosenshine, B. (1979). Content, time, and direct instruction. In P. Peterson & H. J. Walberg (Hrsg.), *Research on teaching: Concepts, findings, and implications* (S. 28-56). Berkeley: McCutchan.
- Sammons, P., Hillman, J., & Mortimore, P. (1995). *Key characteristics of effective schools: A review of school effectiveness research. A report by the Institute of Education for the Office for Standards in Education*. London: Institute of Education.
- Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods*, 7(2), 147-177.
- Scheerens, J., & Bosker, R. J. (1997). *The foundations of educational effectiveness*. Oxford: Pergamon Press.
- Seidel, T., & Shavelson, R. J. (2007). Teaching effectiveness research in the last decade: The role of theory and research design in disentangling meta-analysis results. *Review of Educational Research*, 77, 454-499.
- Senkbeil, M. (2006). Die Bedeutung schulischer Faktoren für die Kompetenzentwicklung in Mathematik und Naturwissenschaften. In Deutsches PISA Konsortium (Hrsg.), *PISA 2003. Untersuchungen zur Kompetenzentwicklung im Verlauf eines Schuljahres* (S. 277-308). Münster: Waxmann.
- Senkbeil, M., Drechsel, B., Rolff, H.-G., Bensen, M., Zimmer, K., Lehmann, R. H., & Neumann, A. (2004). Merkmale und Wahrnehmungen von Schule und Unterricht. In Deutsches PISA Konsortium (Hrsg.), *PISA 2003. Der Bildungsstand der Jugendlichen in Deutschland – Ergebnisse des zweiten internationalen Vergleichs* (S. 296-314). Münster: Waxmann.
- Senkbeil, M., Drechsel, B., & Schöps, K. (2007). Schulische Rahmenbedingungen und Lerngelegenheiten für die Naturwissenschaften. In Deutsches PISA Konsortium (Hrsg.), *PISA 2006. Die Ergebnisse der dritten internationalen Vergleichsstudie* (S. 181-201). Münster: Waxmann.
- Slavin, R. E. (1996). *Education for all*. Lisse: Swets & Zeitlinger.
- Snijders, T. A. B., & Bosker, R. J. (2012). *Multilevel analysis*. London: Sage.
- Spieß, E. (2004). Kooperation und Konflikt. In H. Schuler (Hrsg.), *Organisationspsychologie – Gruppe und Organisation* (S. 193-247). Göttingen: Hogrefe.
- Stanat, P., Rauch, D., & Segeritz, M. (2010). Schülerinnen und Schüler mit Migrationshintergrund. In E. Klieme, C. Artelt, J. Hartig, N. Jude, O. Köller, M. Prenzel, W. Schneider & P. Stanat (Hrsg.), *PISA 2009. Bilanz nach einem Jahrzehnt* (S. 200-230). Münster: Waxmann.

- Steinert, B., Hartig, J., & Klieme, E. (2008). Institutionelle Bedingungen sprachlicher Kompetenzen. In DESI-Konsortium (Hrsg.), *Unterricht und Kompetenzerwerb in Deutsch und Englisch. Ergebnisse der DESI-Studie* (S. 411-450). Weinheim/Basel: Beltz.
- Steinert, B., Klieme, E., Maag Merki, K., Döbrich, P., Halbheer, U., & Kunz, A. (2006). Lehrerkooperation in der Schule. Konzeption, Erfassung, Ergebnisse. *Zeitschrift für Pädagogik*, 52(2), 185-204.
- Stringfield, S. (1994). A model of elementary school effects. In D. Reynolds, B. P. M. Creemers, P. S. Nesselrodt, C. Teddlie, E. C. Shaffer & S. Stringfield (Hrsg.), *Advances in school effectiveness research and practice* (S. 153-187). Oxford: Pergamon.
- Teddlie, C., & Reynolds, D. (Hrsg.) (2000). *The international handbook of school effectiveness research*. London: Falmer Press.
- Wang, M., Haertel, G. D., & Walberg, H. J. (1993). Toward a knowledge base for school learning. *Review of Educational Research*, 63(3), 249-294.
- Weiß, M., & Steinert, B. (2001). Institutionelle Vorgaben und ihre aktive Ausgestaltung – Die Perspektive der deutschen Schulleitungen. In Deutsches PISA-Konsortium (Hrsg.), *PISA 2000: Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich* (S. 427-454). Opladen: Leske + Budrich.

Anschrift der Autorinnen/des Autors

Dr. Brigitte Steinert, Deutsches Institut für Internationale Pädagogische Forschung (DIPF),
Schloßstraße 29, 60486 Frankfurt am Main, Deutschland
E-Mail: steinert@dipf.de

Dr. Jan Hochweber, Deutsches Institut für Internationale Pädagogische Forschung (DIPF),
Schloßstraße 29, 60486 Frankfurt am Main, Deutschland
E-Mail: hochweber@dipf.de

JunProf. Dr. Silke Hertel, Deutsches Institut für Internationale Pädagogische Forschung (DIPF),
Schloßstraße 29, 60486 Frankfurt am Main, Deutschland
E-Mail: hertel@dipf.de

Elternberatung an Schulen im Sekundarbereich

Schulische Rahmenbedingungen, Beratungsangebote der Lehrkräfte und Nutzung von Beratung durch die Eltern

1. Forschungshintergrund

Die Elternberatung ist ein wichtiger Aspekt der Zusammenarbeit von Eltern und Lehrkräften und fest in den Schulalltag in Deutschland integriert. Bereits 1970 wies der Deutsche Bildungsrat auf den hohen Stellenwert von Beratung im Deutschen Schulsystem hin. Die Beratung umfasst dabei ein breites Informations-, Kommunikations- und Unterstützungsangebot von Lehrkräften an Eltern (KMK, 2004). Sie lässt sich theoretisch in Epsteins Rahmenmodell der Beteiligung von Eltern an schulischen Bildungsprozessen verorten (Epstein et al., 2002).

Epstein et al. (2002) unterscheiden sechs Formen der Beteiligung von Eltern: Elterliche Fürsorge und Erziehung, Kommunikation mit der Schule, Beteiligung an freiwilligen/ehrenamtlichen Aktivitäten, Gestaltung von Lernumgebungen im Elternhaus, Mitgestaltung der Schulpolitik im Rahmen von Elternorgans sowie Zusammenarbeit mit der Gemeinde bzw. öffentlichen Einrichtungen. Die Beratung von Eltern durch Lehrkräfte fällt in den Bereich der Kommunikation zwischen Familie und Schule. Dieser umfasst unter anderem regelmäßige Eltern-Lehrer-Gespräche zum Leistungsstand des Kindes und zu Möglichkeiten der Leistungsverbesserung sowie die Information über Schulwahl, Unterrichtsangebote und Aktivitäten an der Schule, Schulpolitik und Reformen. Diese Informationen können auch über zusätzliche Kommunikationsformen (z.B. Anrufe, Elternbriefe, Newsletter und Ähnliches) an die Eltern herangetragen werden.

Der vorliegende Beitrag behandelt mit der Elternberatung somit einen spezifischen Aspekt der elterlichen Beteiligung an Bildungsprozessen. Untersucht werden die Ressourcen und Rahmenbedingungen für Elternberatung, die Beratungsangebote der Lehrkräfte sowie ihre Nutzung durch die Eltern an Schulen im Sekundarbereich. Im Folgenden werden zentrale theoretische Annahmen und empirische Befunde zu den beiden grundlegenden Bereichen dieses Artikels dargestellt: die Beteiligung von Eltern an schulischen Bildungsprozessen und das Beratungsangebot durch Lehrkräfte. Anschließend wird auf der Grundlage des Angebots-Nutzungs-Modells für den Unterricht von Helmke (2009) ein Angebots-Nutzungs-Modell für Elternberatung im schulischen Kontext vorgeschlagen, welches Aspekte der Schule, der Lehrkräfte und der Eltern integriert.

Die Bedeutung der Beteiligung der Eltern an den schulischen Bildungsprozessen ihrer Kinder wurde in vielen Studien belegt. Eine Metaanalyse von Cox (2005) aus dem amerikanischen Raum zeigt, dass durch eine intensive Kooperation von Eltern und Lehrkräften Schülerleistungen verbessert und unangemessene Verhaltensweisen von

Schülern reduziert werden können. Weitere Studien zeigen, dass spezielle Impulse von Lehrkräften an Eltern deren Motivation und Engagement in schulischen Angelegenheiten erhöhen (Corno, 2000; Hoover-Dempsey, Bassler & Burow, 1995; Hoover-Dempsey & Sandler, 1997; Lengua & McMahon, 2000). Nachdem die Forschung zur elterlichen Beteiligung an Bildungsprozessen zunächst insbesondere auf den Kindergarten- und Grundschulbereich fokussierte, finden sich zunehmend Forschungsarbeiten, die den Schwerpunkt auf die Untersuchung entsprechender Prozesse im Sekundarbereich legen. Ein zentraler Befund dieser Studien ist, dass die elterliche Beteiligung an schulischer Bildung auch im Sekundarbereich relevant ist und mit erwünschten Folgen auf Ebene der Schüler (z.B. weniger Verhaltensprobleme, bessere schulische Leistungen) einhergeht (Catsambis, 1998; Hill et al., 2004; Jeynes, 2007). Zudem zeigte eine Metaanalyse von Jeynes (2007), dass diese Effekte unabhängig vom Migrationshintergrund der Familien bestehen. Er deutet dies als Hinweis darauf, dass die elterliche Beteiligung an schulischen Bildungsprozessen ein Ansatzpunkt ist, um die Leistungsunterschiede zwischen Schülern unterschiedlicher kultureller Herkunft zu reduzieren. Befunde einer längsschnittlichen Untersuchung von Hill et al. (2004) legen allerdings nahe, dass sowohl der Zweck als auch die Wirkung von elterlicher Beteiligung an schulischen Bildungsprozessen in Abhängigkeit vom sozioökonomischen sowie vom kulturellen Hintergrund der Eltern variieren können. Die elterliche Beteiligung in Familien mit niedrigerem sozioökonomischen Status wies in ihrer Studie zwar einen positiven Zusammenhang mit den Bildungsaspirationen der Jugendlichen auf, sie führte jedoch nicht zu einer Verbesserung der Voraussetzungen zum Erreichen der höheren Bildungsziele (z.B. weniger Verhaltensauffälligkeiten, bessere schulische Leistungen). Für Familien mit höherem sozioökonomischen Hintergrund zeigten sich positive Zusammenhänge einer entsprechenden Beteiligung mit der Abnahme von Verhaltensproblemen sowie dem Anstieg von schulischen Leistungen und Bildungsaspirationen. Zudem erwies sich die elterliche Beteiligung an Bildungsprozessen für Schüler mit Migrationshintergrund bedeutsamer als für Schüler ohne Migrationshintergrund. Insbesondere für Eltern aus bildungsferneren sozialen Schichten und für Eltern mit Migrationshintergrund scheint die Beteiligung an Bildungsprozessen ihres Kindes einerseits besonders relevant und andererseits eine besondere Herausforderung zu sein (Hill et al., 2004).

Diese Befunde legen nahe, dass sowohl der sozioökonomische als auch der kulturelle Hintergrund der Familie bei der Beratung von Eltern durch Lehrkräfte berücksichtigt werden sollten. Allerdings zeigen eine Sichtung der Curricula der Lehrerbildung sowie Befragungen von Lehrkräften (z.B. Hertel, 2009; Hertel & Djakovic, 2010; Wild, 2003), dass Beratungskompetenzen in der Lehreraus- und -weiterbildung noch nicht systematisch vermittelt werden.

Hier knüpfen Projekte und Studien an, welche in den letzten Jahren entstanden sind und sich mit der Entwicklung von (1) Kompetenzmodellen zur Beratungskompetenz (z.B. Bruder, 2011; Hertel, 2009; Schmitz, Bruder & Hertel, 2011), (2) Instrumenten zur Diagnostik der Beratungskompetenz (z.B. Bruder, 2011; Hertel, 2009; Bruder, Keller, Klug & Schmitz, 2011) sowie (3) Trainingsprogrammen zur Förderung der Beratungskompetenz (z.B. Aich, 2006; Hertel, 2009) befassen. Die bisherigen im deutschen Bil-

derungssystem verorteten Studien fokussieren demnach stark auf die erforderlichen Kompetenzen von Lehrkräften bei der Beratung von Eltern sowie Schülern und ihrer Messung und Förderung.

Bislang liegen keine systematischen Untersuchungen dazu vor, welche Rahmenbedingungen an Schulen für die Beratung von Eltern vorliegen und in welchem Umfang Lehrkräfte Beratungsangebote unterbreiten. Dabei sind Beratungsangebote am Elternsprechtag von Beratungsangeboten zusätzlich zum Elternsprechtag zu unterscheiden. Der Elternsprechtag ist eine strukturelle, durch das Kultusministerium der Bundesrepublik Deutschland und die Schule vorgegebene Form des Beratungsangebots für Eltern. Elterngespräche, die nicht am Elternsprechtag stattfinden, sind häufig individuelle Angebote der Lehrkräfte an die Eltern. Die Anregung und Durchführung solcher Gespräche liegt dann in der Verantwortung der Lehrkräfte. Folglich besteht eine vergleichsweise hohe Autonomie bezogen auf Umfang und Ausgestaltung des Beratungsangebots durch Lehrkräfte zusätzlich zum Elternsprechtag. Befunde einer Studie von Wild (2003) weisen darauf hin, dass Lehrkräfte seltener Beratungsgespräche anbieten, wenn sie sich weniger kompetent fühlen und Elterngesprächen einen geringeren Nutzen zuschreiben. Es ist anzunehmen, dass zusätzlich zu diesen individuellen Faktoren auch die schulischen Rahmenbedingungen für Elternberatung – z.B. die Raumsituation, das Vorhandensein eines Beratungskonzepts, der Stellenwert von Beratung an der Schule – einen Einfluss auf das Beratungsangebot nehmen.

Zur Integration der Befunde zur elterlichen Beteiligung an schulischen Bildungsprozessen sowie der Lehrer- und Schulvariablen, die in einem Zusammenhang mit den Beratungsangeboten für Eltern stehen, bietet sich ein Angebots-Nutzungs-Modell an. Helmke (2009) beschreibt ein solches Modell für den Unterricht. Der Unterricht stellt hierbei eine Lerngelegenheit dar, die von den Schülern genutzt wird. Sowohl die Gestaltung des Unterrichts als auch die Nutzung durch die Schüler wird unter anderem durch kontextuelle Faktoren wie die Schulart oder die Zusammensetzung der Schülerschaft beeinflusst. Schließlich wird durch die Nutzung der Lerngelegenheiten durch die Schüler ein Ertrag – z.B. ein fachlicher oder fächerübergreifender Kompetenzgewinn – erzielt.

Bezogen auf die Elternberatung lässt sich ein entsprechendes Angebots-Nutzungs-Modell wie folgt spezifizieren (siehe Abbildung 1): Das Beratungsangebot wird sowohl von schulischen Struktur- und Prozessmerkmalen als auch von individuellen Merkmalen der Lehrkräfte bestimmt. Strukturmerkmale der Schule umfassen strukturelle Ressourcen – etwa das Vorhandensein eines Besprechungszimmers für Elterngespräche oder das Vorhandensein eines Beratungskonzepts – sowie die Zusammensetzung der Schülerschaft und die Schulart. Als Prozessmerkmal ist der Stellenwert der Elternarbeit an der Schule anzuführen. Auf der Ebene der Lehrkraft können individuelle Kompetenzüberzeugungen in Bezug auf die Beratung von Eltern, das professionelle Rollenverständnis sowie der wahrgenommene Nutzen von Elterngesprächen einen Einfluss auf das Beratungsangebot nehmen. Bezug nehmend auf die Befunde zur Bedeutung von sozioökonomischem und kulturellem Hintergrund der Eltern bei der Beteiligung an schulischen Bildungsprozessen sollte auch die Zusammensetzung der Schülerschaft einen

Einfluss auf die Beratungsangebote nehmen. D.h. die Schule bzw. die Lehrkräfte sollten das Beratungsangebot auch unter Berücksichtigung der sozioökonomischen und der kulturellen Herkunft der Eltern gestalten und somit adaptiv auf die spezifischen Bedarfe dieser Eltern reagieren. Weiterhin ist zu vermuten, dass die Nutzung der Beratungsangebote durch die Eltern durch strukturelle Familienmerkmale, beispielsweise den sozioökonomischen Hintergrund, sowie Prozessmerkmale, etwa die Förderung des Kindes durch die Eltern oder der Leistungsstand des Kindes, beeinflusst wird. Durch die Nutzung von Beratungsangeboten können sich Veränderungen in der Unterstützung und Förderung des Kindes durch die Eltern ergeben, welche in der Konsequenz zu einem günstigeren Lernverhalten, besseren schulischen Leistungen und höheren Bildungsaspirationen bei den Schülern führen.

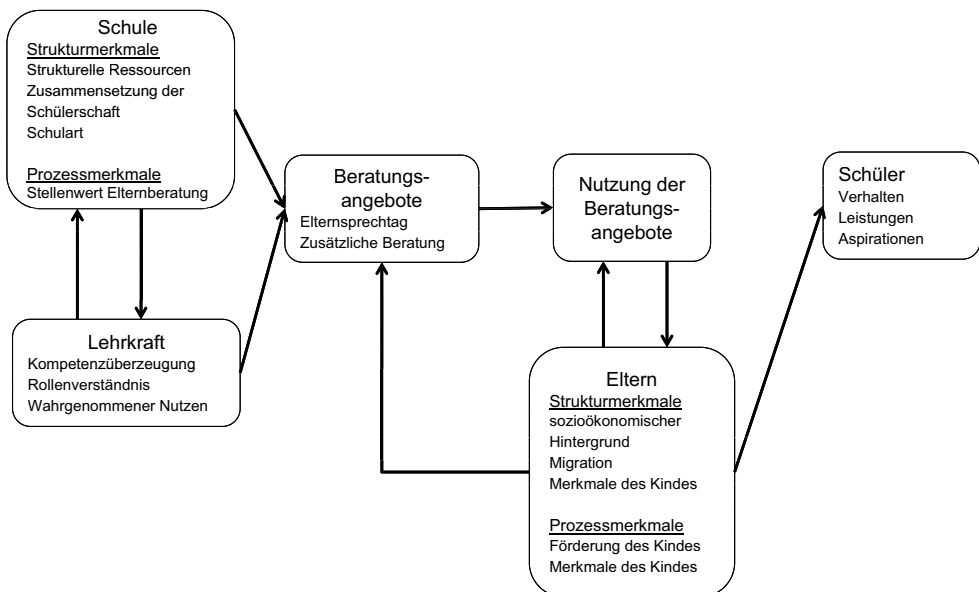


Abb. 1: Angebots-Nutzungs-Modell für die Elternberatung im schulischen Kontext

Im Rahmen dieses Beitrags sollen die zentralen Aspekte dieses Angebots-Nutzungs-Modells systematisch beschrieben und in Bezug auf spezifische Zusammenhänge überprüft werden. Dabei können nur Aussagen über die Angebote und deren Nutzung getroffen werden. Analysen zum Ertrag der Nutzung von Elternberatung – d.h. zu Veränderungen auf Ebene der Schüler – können nicht erfolgen. Entsprechende Analysen können nur auf der Grundlage eines längsschnittlichen Designs vorgenommen werden. Für die Auswertungen liegen allerdings nur Querschnittsdaten vor.

2. Fragestellung

In dem vorliegenden Beitrag werden vier Fragestellungen untersucht:

- (1) Welche Ausprägungen der prozessbezogenen und strukturellen Rahmenbedingungen für die Elternberatung liegen an den Schulen vor, wie häufig werden Beratungsangebote unterbreitet und wie stark ist die Nachfrage von Eltern nach solchen Angeboten?
- (2) Beeinflussen die Zusammensetzung der Schülerschaft in Bezug auf Leistung, sozioökonomischen Hintergrund und Migrationshintergrund sowie die Schularzt Angebot und Nachfrage von Beratung sowie die strukturellen Ressourcen für die Elternberatung an der Schule?
- (3) In welchem Umfang unterbreiten Lehrkräfte Beratungsangebote und in welchem Zusammenhang mit individuellen sowie schulischen Merkmalen stehen die Beratungsangebote?
- (4) Zu welchen Themen wünschen sich Eltern Beratung durch Lehrkräfte und mit welchen strukturellen und prozessbezogenen Merkmalen der Familie geht die Nutzung von Beratungs- und Informationsangeboten durch die Eltern einher?

3. Daten und Methode

Um die oben genannten Forschungsfragen beantworten zu können, wurden Schulleiter, Lehrkräfte des Fachkollegiums Deutsch sowie Eltern mittels Fragebögen befragt. Die Befragung war Teil der Erhebung des Programmes for International Student Assessment (PISA) in Deutschland im Jahr 2009. Die internationalen PISA-Fragebögen wurden erweitert und im nationalen Fragebogenabschnitt um den thematischen Bereich der Elternberatung ergänzt.

3.1 Stichprobe

Die vorliegenden Auswertungen beziehen sich auf die Angaben von 213 Schulleitern (weiblich = 25.7%), 2201 Lehrkräften des Fachkollegiums Deutsch (weiblich = 70.8%) der PISA-2009-Schulen (inkl. Berufsschulen und Förder- bzw. Sonderschulen) sowie 6872 Eltern von Schülern in der 9. Jahrgangsstufe (national erweiterte Stichprobe). Die Teilnahmequoten betrugen 94.2 Prozent auf Schulleiterebene, 76.7 Prozent auf Ebene der Lehrkräfte sowie 72.6 Prozent auf Elternebene.

Die 213 Schulleiter verteilten sich auf 40 Hauptschulen (18.8%), 15 Schulen mit mehreren Bildungsgängen (7.0%), 55 Realschulen (25.8%), 18 Integrierte Gesamtschulen (8.5%), 63 Gymnasien (29.6%) sowie neun Berufsschulen (4.2%) und 13 Sonder- und Förderschulen (6.1%). 519 Lehrkräfte (24.7%) unterrichteten zum Befragungszeitpunkt Deutsch in der 9. Jahrgangsstufe. 913 Eltern waren nicht-deutscher Her-

kunft (13.7%), der mittlere HISEI lag bei 48.7 ($SE = .55$, $SD = 15.51$, Range 16–90). Die Korrelation zwischen HISEI und Migrationshintergrund war statistisch signifikant ($p < .001$) und betrug $r = -.26$ ($SE = .02$).

Repräsentative Aussagen für Deutschland sind nur auf Schulebene möglich, hier war die Teilnahme an der Befragung verpflichtend. Lehrkräfte und Eltern beantworteten die Fragebögen auf freiwilliger Basis, sodass von Selektionseffekten auszugehen ist (vgl. auch Hertel, Hochweber, Steinert & Klieme, 2010; Hertel, Jude & Naumann, 2010). Für den Elternfragebogen zeigen sich Ausfälle insbesondere bei Eltern mit Migrationshintergrund. Eltern mit höherem sozioökonomischen Status sowie Eltern von Kindern an Realschulen bzw. an Gymnasien nehmen tendenziell häufiger an der Befragung teil (vgl. Hertel, Jude & Naumann, 2010).

3.2 Instrumente

Die Fragebogenitems wurden aus nationalen Studien zur Elternberatung von Hertel (2009) sowie von Arbeiten im Rahmen des DFG-Projekts *Beratungskompetenz* im Schwerpunktprogramm 1293 (Schmitz et al., 2011) übernommen bzw. in Anlehnung an diese entwickelt. Die Skalierung der Fragebogendaten erfolgte auf der Grundlage der Klassischen Testtheorie (Hertel, Steinert, Mildner & Rauch, in Vorb.).

Die Schulleiter machten Angaben zu den *Ressourcen an der Schule im Hinblick auf Elternarbeit* (5 Items, Beispielitem: An unserer Schule verfügen wir über ein Besprechungszimmer, das wir für Elterngespräche nutzen können), zum *Stellenwert von Elternarbeit an der Schule* (1 Item: Elternarbeit hat einen hohen Stellenwert an unserer Schule) sowie zu Beratungsangeboten und Elternnachfrage für Beratung an der Schule (4 Items, Beispielitem: Die Eltern unserer Schülerinnen/Schüler äußern nur selten den Wunsch, von Lehrkräften beraten zu werden). Die Fragen konnten auf einer Skala von (1) *stimmt überhaupt nicht* bis (4) *stimmt ganz genau* beantwortet werden.

Die Lehrkräfte wurden nach der *Anzahl der Elterngespräche* gefragt, die sie am Elternsprechtag sowie zusätzlich zum Elternsprechtag durchführen. Zur Beantwortung dieser Fragen konnten sie aus sieben Kategorien von (1) *An meiner Schule gibt es keinen Elternsprechtag* bzw. *Ich führe Elterngespräche nur am Elternsprechtag* bis (7) *Über 20 Gespräche* wählen. Zudem wurden ihre *Kompetenzüberzeugungen in Bezug auf Beratung* (30 Items, Cronbachs $\alpha = .90$, Beispielitem: In Gesprächen mit Eltern strukturiere ich die Aussagen der Eltern, fasse sie zusammen und gebe sie in meinen eigenen Worten wieder), ihre *Überzeugungen zur Lehrerrolle in Bezug auf Beratung* (3 Items, Cronbachs $\alpha = .80$, Beispielitem: Elternberatung ist ein wichtiger Bestandteil des Lehrerberufs) sowie ihre *Überzeugungen zum Nutzen von Elterngesprächen* (3 Items, Cronbachs $\alpha = .72$, Beispielitem: Der Aufwand, den eine intensive Elternberatung mit sich bringt, steht in keinem Verhältnis zu dem Ergebnis, das erzielt werden kann) erfasst. Schließlich beurteilten die Lehrkräfte – wie die Schulleiter – die Rahmenbedingungen für die Elternarbeit an der Schule; für die Skalenbildung auf Lehrer-

ebene wurden Prozess- und Strukturmerkmale zusammengefasst (6 Items, Cronbachs $\alpha = .71$). Diese Fragen konnten auf einer Skala von (1) *stimmt überhaupt nicht* bis (4) *stimmt ganz genau* beantwortet werden.

Die Nutzung der Beratungs- und Informationsangebote der Lehrkräfte an den Schulen durch die Eltern wurde mit der Frage, ob die Eltern im letzten Schuljahr an einem Gesprächstermin oder einem Elternabend an der Schule teilgenommen haben, erfasst. Diese Frage war eine nationale Ergänzung im internationalen Fragebogen für Eltern und konnte mit *ja* oder *nein* beantwortet werden. Zusätzlich werden in diesem Artikel die Angaben der Eltern zu der Förderung ihres Kindes im Grundschulalter (z.B. Bücher lesen, Geschichten erzählen, Lieder singen, Buchstaben oder Wörter schreiben, 9 Items, Cronbachs $\alpha = .78$) sowie ihre aktuelle Förderung des Kindes (z.B. über politische oder soziale Themen diskutieren, darüber sprechen, was das Kind gerade liest, gemeinsam etwas am PC tun, 9 Items, Cronbachs $\alpha = .66$) erfasst. Beide Skalen stammen aus dem internationalen Elternfragebogen und konnten auf einer vierstufigen Skala von (1) *nie oder fast nie* bis (4) *jeden Tag oder fast jeden Tag* beantwortet werden. National ergänzt wurde zudem die Frage nach Themenbereichen, für welche Eltern Lehrkräfte als wichtige Ansprechpersonen ansehen. Es wurden acht Inhaltsbereiche vorgegeben (z.B. Fragen zum Leistungsstand und zur Schulwahl, Fragen zur Versetzung und zum Schulwechsel, Fragen zu Hilfestellungen bei den Hausaufgaben), die Eltern konnten auf einer Skala von (1) *stimmt überhaupt nicht* bis (4) *stimmt ganz genau* antworten.

Die soziale Herkunft der Familie wurde mit dem internationalen sozioökonomischen Index (ISEI) bestimmt. Der Index wird auf der Grundlage der Berufe der Mutter und des Vaters gebildet. Der Wertebereich liegt zwischen 16 (z.B. ungelernte Hilfskraft) und 90 (z.B. RichterIn/Richter). Für die Bestimmung des HISEI (highest ISEI) werden die Werte von Mutter und Vater verglichen, der höhere der beiden Werte wird herangezogen. Die Bildung des HISEI ist seit PISA 2000 unverändert (vgl. Ehmke & Jude, 2010).

Der Migrationshintergrund der Familie wurde über die Angaben der Schüler zu ihrer Herkunft sowie zur Herkunft ihrer Eltern bestimmt. Zur Definition des Migrationshintergrunds wurde die OECD-Zuordnung herangezogen. Schüler, bei denen beide Elternteile im Ausland geboren sind, wurden zur Gruppe der Jugendlichen mit Migrationshintergrund gezählt (vgl. Stanat, Rauch & Segeritz, 2010). Dementsprechend wurden auch ihre Eltern der Gruppe von Eltern mit Migrationshintergrund zugeordnet. Bei den Analysen erfolgte keine Unterscheidung zwischen der 1. Generation (beide Elternteile und Schülerin/Schüler im Ausland geboren) und der 2. Generation (beide Elternteile im Ausland geboren, Schülerin/Schüler in Deutschland geboren).

Als Indikatoren für die Schülerleistung wurden die Leistungstests für die Domänen Lesen, Mathematik und Naturwissenschaften aus PISA 2009 herangezogen (siehe Frey, Heinze, Mildner, Hochweber & Asseburg, 2010; Naumann, Artelt, Schneider & Stanat, 2010; Rönnebeck, Schöps, Prenzel, Mildner & Hochweber, 2010).

3.3 Datenanalyse

Zur Beschreibung der Rahmenbedingungen für Elternberatung an Schulen (Fragestellung 1) wurden deskriptive Analysen mittels SPSS (Version 20) durchgeführt. Die Datenbasis bildeten die Angaben der Schulleiter in der PISA-2009-Stichprobe, die mit dem Schulgewicht gewichtet wurden. Für die deskriptiven Auswertungen wurden die Angaben der Schulleiter dichotomisiert (*stimmt gar nicht/stimmt eher nicht* vs. *stimmt eher/stimmt ganz genau*). Die Analysen wurden sowohl für die Gesamtstichprobe als auch aufgeschlüsselt nach Schularten durchgeführt. Es wurden alle in PISA 2009 untersuchten Schularten – Sonder- und Förderschulen, Hauptschulen, Schulen mit mehreren Bildungsgängen, Realschulen, Integrierte Gesamtschulen, Gymnasien und Berufsschulen – einbezogen.

Zur Beantwortung der drei weiteren Fragestellungen wurden multivariate logistische Regressionsanalysen durchgeführt. Datenbasis waren die Angaben der Schulleiter, Lehrkräfte und Eltern von Kindern an allgemeinbildenden Schulen (ohne Sonder- und Förderschulen sowie Berufsschulen). Zusätzlich wurden die auf Schulebene aggregierte mittlere Lesekompetenz und der mittlere sozioökonomische Hintergrund der Schüler sowie der Anteil von Schülern mit Migrationshintergrund in der Schülerschaft herangezogen. Fehlende Werte für den sozioökonomischen Hintergrund der Schüler wurden imputiert. Die Leistungswerte der Schüler im Bereich der Lesekompetenz wurden auf der Basis von Plausible Values (OECD, 2011) in die Analysen einbezogen. Entsprechend wurden fünf Datensätze erzeugt, die parallel in die Analysen eingingen. Die Schulart wurde bei den Analysen in der für PISA üblichen Dummy-Kodierung berücksichtigt. Schulen mit Hauptschulbildungsgang und Gymnasien werden mit der Referenzgruppe Realschulen verglichen, Berufsschulen sowie Sonder- und Förderschulen wurden nicht einbezogen. Die Analysen erfolgten mittels Mplus 5.1 (Muthén & Muthén, 1998-2005). Die hierarchisch geschachtelte Datenstruktur wurde über die Type=Complex-Option berücksichtigt; als Schätzer wurde der MLR-Schätzer (maximum likelihood estimation with robust standard errors) verwendet. Die Signifikanzprüfung erfolgte zweiseitig. Für kontinuierliche Prädiktoren werden als standardisierte Koeffizienten Std YX-Werte angegeben, für dichotome Prädiktoren werden die Std-Koeffizienten berichtet.

Um zu untersuchen, welche Faktoren einen Zusammenhang mit Angebot und Nachfrage von Beratung sowie mit den schulischen Ressourcen für Elternberatung aufweisen (Fragestellung 2), wurden als Kriterium die jeweiligen Ressourcen für Elternarbeit bzw. die Beratungsangebote der Schule und deren Nachfrage (dichotome Ausprägung: vorhanden/nicht vorhanden) herangezogen. Als Prädiktoren wurden Prozessvariablen und Strukturvariablen an der Schule eingeführt. Die Angaben der Schulleiter sowie die auf Schulebene aggregierten Daten (mittlere Lesekompetenz der Schüler, mittlerer sozioökonomischer Status der Schüler, Anteil von Schülern mit Migrationshintergrund) wurden mit dem Schulgewicht gewichtet in die Analysen einbezogen.

Die dritte Fragestellung zielt auf die Zusammenhänge von individuellen und schulischen Faktoren mit dem Beratungsangebot von Lehrkräften ab. Zur Beantwortung dieser Fragestellung wurden logistische Regressionsanalysen für ordinalskalierte Variablen durchgeführt. Die Datenbasis bildeten die Angaben von Lehrkräften sowie die auf

Schulebene aggregierte mittlere Lesekompetenz der Schüler, der mittlere sozioökonomische Status der Schüler sowie der Anteil von Schülern mit Migrationshintergrund. Für die Analysen zum Beratungsangebot am Elternsprechtag wurden nur Lehrkräfte von Schulen einbezogen, an denen es einen Elternsprechtag gibt. Bei beiden abhängigen Variablen (Beratungsangebot am Elternsprechtag, Beratungsangebot zusätzlich zum Elternsprechtag) wurden Antwortkategorie 2 (weniger als 3 Gespräche) und Antwortkategorie 3 (3–5 Gespräche) zusammengefasst. Bei den Auswertungen zu den Angaben der Lehrkräfte wurde auf eine Gewichtung verzichtet, da die Zielpopulation der Lehrkräfte des Fachkollegiums Deutsch für die Beratungsangebote an Schulen keine repräsentative Population darstellt.

Die vierte Fragestellung bezieht sich auf die Beschreibung des Beratungswunsches der Eltern sowie auf die Nutzung der Beratungsangebote an Schulen durch die Eltern. Zur Beschreibung des Beratungswunsches der Eltern wurden deskriptive Analysen mit dem SPSS-Macro für die Auswertung von PISA-2009-Daten durchgeführt (OECD, 2011). Für die Auswertungen zur Nutzung der Beratungsangebote wurden logistische Regressionsanalysen durchgeführt. Das Kriterium war die Teilnahme an Elternabenden bzw. Beratungsgesprächen der Eltern (dichotom: ja/nein). Es wurden zwei Modelle berechnet: In Modell I werden nur Schülermerkmale als Prädiktoren berücksichtigt. In Modell II werden zusätzlich zu den Schülermerkmalen Struktur- und Prozessmerkmale der Familie eingeführt. Die Angaben der Eltern wurden mit dem Schülergewicht gewichtet in die Datenauswertungen einbezogen. Die berichteten Prozentwerte lassen sich als prozentualer Anteil der Schüler in der 9. Jahrgangsstufe, deren Eltern die dargestellte Meinung vertreten bzw. die Beratungs- und Informationsangebote nutzen, interpretieren.

4. Ergebnisse

Im Folgenden werden die Ergebnisse der Auswertungen zu den vier Fragestellungen dargestellt. Diese beziehen sich (1) auf die Rahmenbedingungen für die Beratung von Eltern an Schulen, (2) die Zusammenhänge der Rahmenbedingungen für Elternberatung mit den Struktur- und Prozessmerkmalen der Schule, (3) die Einflussfaktoren auf Beratungsangebote von Lehrkräften sowie auf (4) die Nutzung von Beratungs- und Informationsangeboten durch die Eltern.

4.1 *Rahmenbedingungen für die Beratung von Eltern an Schulen im Sekundarbereich*

Die erste Fragestellung des vorliegenden Artikels untersucht, (1) welche Rahmenbedingungen aus Sicht der Schulleiter für die Beratung von Eltern vorliegen und (2) wie stark das Beratungsangebot und Nachfrage nach Beratung durch die Eltern an den Schulen ausgeprägt sind.

Der Stellenwert von Elternarbeit an der Schule wird von den Schulleitern mehrheitlich als sehr hoch eingeschätzt (*stimmt eher/stimmt genau*: 89.6%). Dabei wird an den meisten Schulen Elternarbeit auch zusätzlich zum Elternsprechtage durchgeführt (76.7%). Dementsprechend gibt ein Großteil der Schulleiter an, dass es an ihrer Schule vielfältige Beratungsangebote für Eltern gibt (75.4%). Nur an wenigen Schulen werden Beratungsgespräche lediglich dann geführt, wenn es „brennt“ (10.1%). An etwa zwei Dritteln der Schulen wird von den Schulleitern zudem eine deutliche Nachfrage der Eltern nach Beratung wahrgenommen (66.7%). Diese Befunde unterstreichen den Stellenwert der Beratung im Schulalltag. Doch welche strukturellen Ressourcen stehen für Beratungsgespräche zur Verfügung?

Ein Großteil der Schulen verfügt nach Angabe der Schulleiter über ein Besprechungszimmer für Beratungsgespräche (88.1%). Ein Beratungskonzept für die Elternarbeit (56.9%) bzw. spezifisch für den Bereich der Lernförderung (50.3%) liegt jeweils an etwa der Hälfte der Schulen vor. Die Unterstützung bei Beratungsaufgaben ist nach Angaben der Schulleiter deutlich ausgeprägt: Es besteht eine hohe Kooperation im Kollegium in Bezug auf Beratungsgespräche (83.2%), gut zwei Drittel der Schulen (70.5%)

Rahmenbedingungen	Gesamt	Hauptschule	Schule mit mehreren Bildungsgängen	Realschule	Integrierte Gesamtschule	Gymnasium	Berufsschule	Sonderschule
	% (SE)	% (SE)	% (SE)	% (SE)	% (SE)	% (SE)	% (SE)	% (SE)
Stellenwert von Elternarbeit	89.6 (.00)	90.4 (.00)	96.2 (.01)	90.1 (.01)	96.3 (.01)	87.5 (.01)	59.0 (.02)	100.0 (.00)
Elternarbeit überw. am Elternsprechtage	23.3 (.00)	25.6 (.01)	36.8 (.01)	21.6 (.01)	17.0 (.02)	23.9 (.01)	32.3 (.02)	7.3 (.01)
Vielfältige Beratungsangebote	75.4 (.00)	77.7 (.01)	87.3 (.01)	63.1 (.01)	96.9 (.01)	66.6 (.01)	62.3 (.02)	92.7 (.01)
Elterngespräche nur wenn es „brennt“	10.1 (.00)	7.5 (.00)	9.0 (.00)	13.9 (.01)	7.1 (.01)	12.5 (.01)	37.7 (.02)	0.0 (.00)
Nachfrage der Eltern	66.7 (.00)	38.3 (.01)	60.0 (.01)	19.1 (.01)	9.6 (.01)	23.0 (.01)	100.0 (.00)	10.2 (.01)
Besprechungszimmer	88.1 (.00)	78.7 (.01)	84.5 (.01)	91.0 (.01)	84.7 (.02)	89.5 (.01)	95.7 (.01)	100.0 (.00)
Beratungskonzept Elternarbeit	56.9 (.00)	44.5 (.01)	60.0 (.01)	58.7 (.01)	85.3 (.02)	54.7 (.01)	38.8 (.02)	80.4 (.01)
Beratungskonzept Lernförderung	50.3 (.00)	34.9 (.01)	64.7 (.01)	44.0 (.01)	72.7 (.02)	49.9 (.01)	38.8 (.02)	78.2 (.01)
Unterstützung im Kollegium	83.2 (.00)	92.3 (.00)	80.0 (.01)	86.9 (.01)	100.0 (.00)	73.8 (.01)	44.3 (.02)	92.7 (.01)
Experten-Netzwerk	70.5 (.00)	76.0 (.01)	23.7 (.01)	62.5 (.01)	96.9 (.01)	76.6 (.01)	95.7 (.01)	87.8 (.01)

Anmerkung: Angegebene Prozentwerte beziehen sich auf Zustimmung (stimmt eher/stimmt genau) der Schulleiter

Tab. 1: Stellenwert, Beratungsangebote, Beratungsnachfrage und strukturelle Rahmenbedingungen für Elternberatung an Schulen im Sekundarbereich

verfügen über ein Netzwerk von Experten (z.B. Schulpsychologen oder Beratungsstellen), auf welches in konkreten Beratungssituationen zurückgegriffen werden kann. Die Ergebnisse sind in Tabelle 1 dargestellt.

Die differenzierte Betrachtung nach Schularten zeigt, dass die Elternarbeit in den meisten Schularten einen hohen Stellenwert aufweist. Die Zustimmung reicht von 100% bei Schulleitern an Sonder- und Förderschulen bis 87.5% an Gymnasien. Eine Ausnahme bilden dabei die Berufsschulen, hier ist der Stellenwert von Elternarbeit mit einer Zustimmung von 59.0% nach Ansicht der Schulleiter deutlich geringer.

In der Gesamtbetrachtung zeigt sich, dass an Integrierten Gesamtschulen im Vergleich zu den anderen Schularten die günstigsten Rahmenbedingungen für Elternberatung vorliegen: Es bestehen vielfältige Beratungsangebote für Eltern, eine sehr hohe Unterstützung im Kollegium und eine starke Einbindung in Expertennetzwerke. Zudem liegen sehr häufig Beratungskonzepte für die Elternarbeit bzw. spezifisch für die Lernförderung vor. Vergleichbar günstige Rahmenbedingungen finden sich sonst nur an Sonder- und Förderschulen. Zudem zeigt sich, dass Beratungskonzepte für die Elternarbeit bzw. die Lernförderung vergleichsweise selten an Hauptschulen vorliegen.

Doch welche Faktoren stehen in einem Zusammenhang mit dem Angebot und der Nachfrage von Elternberatung sowie den strukturellen Ressourcen für Elterngespräche an Schulen?

4.2 Zusammenhänge von Angebot und Nachfrage von Elternberatung sowie Ressourcen für Beratungsgespräche mit dem Stellenwert von Elternberatung an der Schule und der Komposition der Schülerschaft

Nachdem in dem vorangegangenen Abschnitt die Rahmenbedingungen für die Beratung von Eltern an Schulen im Sekundarbereich beschrieben wurden, wird in diesem Abschnitt untersucht, welche Faktoren einen Zusammenhang mit dem Angebot, der Nachfrage und der Verfügbarkeit von Ressourcen aufweisen. Dabei wird in Anlehnung an das Angebots-Nutzungs-Modell für die Elternberatung in schulischen Kontexten unterschieden zwischen *Prozessvariablen*, spezifisch dem Stellenwert von Elternarbeit an der Schule, sowie *Strukturvariablen*, wie die Zusammensetzung der Schülerschaft hinsichtlich des Kompetenzniveaus, des sozioökonomischen Hintergrunds und des Anteils von Schülern mit Migrationshintergrund sowie der Schulart. Prozessvariablen und Strukturvariablen werden als Voraussetzungen, das heißt als beeinflussende Faktoren im Sinne von Prädiktoren angesehen. Als abhängige Variablen werden (1) für das Beratungsangebot und die Nachfrage von Beratung das *Durchführen von Elternarbeit überwiegend am Elternsprechtag*, die *Vielfältigkeit des Beratungsangebots*, das *Führen von Elterngesprächen nur wenn es brennt* sowie *eine geringe Nachfrage der Eltern nach Beratung* herangezogen. Zudem werden (2) bezogen auf die strukturellen Ressourcen die *Verfügbarkeit eines Besprechungszimmers für Elternberatung*, das *Vorhandensein eines Beratungskonzepts für Elternarbeit*, die *hohe Unterstützung im Kollegium* sowie die *Einbindung in ein Expertennetzwerk* betrachtet. Die Ergebnisse der Auswertungen sind in Tabelle 2 dargestellt.

Prädiktoren	Angebot und Nachfrage von Elternberatung											
	Beratung über- wiegend am Eltern- sprechtag			Vielfältiges Beratungs- angebot			Beratung nur wenn es „brennt“			Geringe Nachfrage der Eltern		
	<i>b</i>	(SE)	β	<i>b</i>	(SE)	β	<i>b</i>	(SE)	β	<i>b</i>	(SE)	β
<i>Prozess- variablen</i>												
Stellenwert von Eltern- arbeit	-.55	.18	-.31	.76	.23	.42	-.56	.25	-.30	-.49	.21	-.27
<i>Struktur- variablen</i>												
Mittlere Lese- kompetenz	.11	.01	.57	-.01	.01	-.32	.01	.01	.30	.001	.01	.03
Mittlerer sozio- ökonomischer Hintergrund	-.05	.05	-.03	.08	.04	.49	-.10	.10	-.60	-.11	.05	-.63
Anteil von Schülern mit Migrations- hintergrund	1.46	.99	.22	1.47	.88	.21	1.09	1.37	.15	-.02	.91	-.003
Schularten mit Hauptschul- bildungsgang	-.11	.53	-.11	-.13	.55	-.13	.38	.77	.38	1.02	.59	1.02
Gymnasium	.55	.44	.55	.50	.50	.50	-.93	.58	-.93	.22	.50	.22
R ²	.23			.30			.34			.30		

Anmerkung: Fettgedruckte Regressionskoeffizienten sind statistisch signifikant ($p < .05$);
Testung: zweiseitig.

Tab. 2a: Regressionskoeffizienten der Zusammenhänge von Prozessvariablen und Strukturvariablen mit Angebot und Nachfrage von Elternberatung

Über die Analysen für Angebot und Nachfrage sowie für die Ressourcen für Elternarbeit an der Schule hinweg erweist sich die Prozessvariable *Stellenwert von Elternarbeit an der Schule* als zentraler Prädiktor. An Schulen, die der Elternarbeit einen hohen Stellenwert einräumen, findet die Beratung häufiger zusätzlich zum Elternsprechtage statt, es gibt ein vielfältiges Beratungsangebot für Eltern und Beratungsgespräche werden nicht erst dann geführt, wenn „es brennt“. Auch die Nachfrage der Eltern nach Beratung ist an diesen Schulen höher ausgeprägt. Zudem geht ein höherer Stellenwert von Elternarbeit an der Schule mit einer günstigeren Ressourcenausstattung einher. So bestehen an den entsprechenden Schulen häufiger Beratungskonzepte für die Elternarbeit sowie eine höhere Unterstützung im Kollegium.

Prädiktoren	Ressourcen für Elternberatung an der Schule											
	Besprechungs- zimmer			Beratungskonzept für die Elternarbeit			Hohe Unterstützung im Kollegium			Einbindung in ein Expertennetzwerk		
	<i>b</i>	(SE)	β	<i>b</i>	(SE)	β	<i>b</i>	(SE)	β	<i>b</i>	(SE)	β
Prozess- variablen												
Stellenwert von Elternarbeit	.004	.18	.003	.60	.19	.36	.65	.22	.32	.30	.20	.18
Struktur- variablen												
Mittlere Lese- kompetenz	.01	.01	.32	.001	.01	.04	-.01	.01	-.56	-.01	.01	-.68
Mittlerer sozio- ökonomischer Hintergrund	-.03	.05	-.17	-.03	.04	-.21	0.13	.06	.69	.04	.04	.28
Anteil von Schülern mit Migrations- hintergrund	-.51	1.01	-.09	.22	.92	.04	-1.48	.88	-.19	1.06	.91	.17
Schularten mit Hauptschul- bildungsgang	-.29	.65	-.29	.11	.47	.11	-.94	.53	-.94	.64	.47	.64
Gymnasium	-.25	.54	-.25	-.46	.39	-.46	.31	.50	.31	-.51	.45	-.51
R ²	.07			.13			.42			.14		

Anmerkung: Fettgedruckte Regressionskoeffizienten sind statistisch signifikant (p < .05);
Testung: zweiseitig.

Tab. 2b: Regressionskoeffizienten der Zusammenhänge von Prozessvariablen und Struktur-
variablen mit Ressourcen für Elternberatung

Für die schulischen Strukturvariablen ergeben sich nur wenige statistisch signifikante Zusammenhänge mit den abhängigen Variablen. Es zeigt sich, dass Schulen, an denen die mittlere Lesekompetenz der Schüler geringer ausgeprägt ist, weniger stark in Expertennetzwerke eingebunden sind. An Schulen mit einem höheren mittleren sozioökonomischen Hintergrund der Schüler bestehen eine größere Vielfalt an Beratungsangeboten und eine höhere Unterstützung im Kollegium. Für den Anteil an Schülern mit Migrationshintergrund sowie für die Schulart ergeben sich keine statistisch signifikanten Zusammenhänge mit dem Angebot und der Nachfrage von Elternberatung sowie mit den verfügbaren Ressourcen an der Schule.

4.3 *Beratungsangebot von Lehrkräften sowie individuelle und schulische Bedingungsfaktoren*

Beratungsangebote von Lehrkräften

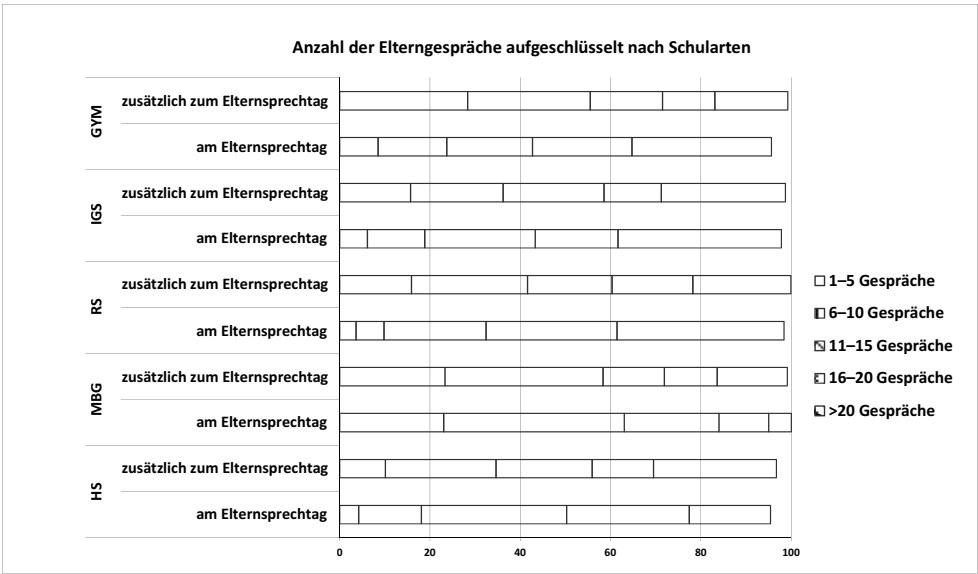
In diesem Abschnitt werden zunächst deskriptive Analysen zu den Häufigkeiten der am Elternsprechtag sowie zusätzlich zum Elternsprechtag durchgeführten Beratungsgespräche durch die Lehrkräfte berichtet. Anschließend werden die Zusammenhänge dieser Beratungsangebote mit individuellen und schulischen Bedingungsfaktoren untersucht.

Über alle Schularten hinweg zeigt sich, dass knapp ein Drittel der Lehrkräfte an einem Elternsprechtag mehr als 20 Elterngespräche führt (28.5%). Jeweils knapp ein Viertel der Lehrkräfte führen zwischen elf und 15 (22.8%) bzw. 16 bis 20 (23.1%) Gespräche mit Eltern. Etwa ein Sechstel der Lehrkräfte aus dem Fachkollegium Deutsch führt sechs bis zehn Elterngespräche an einem Elternsprechtag (14.0%), weniger als fünf Gespräche führen 7.9 Prozent der Lehrkräfte. Nur sehr wenige Lehrkräfte (3.8%) geben an, dass es an ihrer Schule keinen Elternsprechtag gibt. Insgesamt führen etwa drei Viertel der Lehrkräfte (74.4%) mindestens elf Elterngespräche an einem Elternsprechtag.

Für die zusätzlich zum Elternsprechtag durchgeführten Elterngespräche ergibt sich Folgendes: 99.5 Prozent der Lehrkräfte des Fachkollegiums Deutsch führen zusätzlich zum Elternsprechtag Gespräche mit den Eltern ihrer Schüler. Insgesamt führen drei Viertel der Lehrkräfte bis zu 20 Elterngespräche pro Schuljahr zusätzlich zum Elternsprechtag: Etwa ein Viertel der Lehrkräfte führt zusätzlich sechs bis zehn Gespräche pro Schuljahr (26.1%). Jeweils ein Fünftel der Lehrkräfte führt in einem Schuljahr zusätzlich zum Elternsprechtag bis zu fünf (20.4%), elf bis 15 (18.0%) bzw. mehr als 20 Elterngespräche (21.3%). Etwa ein Sechstel der Lehrkräfte führt zwischen 16 und 20 Gespräche mit Eltern außerhalb des Elternsprechtags.

Abbildung 2 zeigt die Angaben der Lehrkräfte zu den Gesprächen am Elternsprechtag sowie zu den zusätzlichen Gesprächen aufgeschlüsselt nach Schularten.

Bei einer differenzierten Betrachtung der Beratungsangebote aufgeschlüsselt nach Schularten zeichnen sich folgende Tendenzen ab: Lehrkräfte an Integrierten Gesamtschulen führen Elterngespräche häufig sowohl am Elternsprechtag als auch zusätzlich zum Elternsprechtag. An Realschulen und Gymnasien bieten Lehrkräfte Elterngespräche vornehmlich am Elternsprechtag an. An Hauptschulen sowie Sonder- und Förderschulen nimmt der Elternsprechtag im Vergleich einen geringeren Stellenwert ein; die Lehrkräfte führen Elterngespräche hier häufiger an zusätzlich vereinbarten Terminen. An Schulen mit mehreren Bildungsgängen bieten die Lehrkräfte verglichen mit Lehrkräften der anderen Schularten seltener Elterngespräche an – dies gilt sowohl für den Elternsprechtag als auch für zusätzlich vereinbarte Gesprächstermine. Doch welche Faktoren beeinflussen das individuelle Beratungsangebot der Lehrkräfte?



Anmerkung: HS = Hauptschule; MBG = Schule mit mehreren Bildungsgängen, RS = Realschule; IGS = Integrierte Gesamtschule, GYM = Gymnasium; Förder- bzw. Sonderschulen und Berufsschulen sind in der Abbildung nicht enthalten; 0.5% der Lehrkräfte geben an, Elterngespräche ausschließlich am Elternsprechtag zu führen.

Abb. 2: Anzahl der Elterngespräche am Elternsprechtag und zusätzlich zum Elternsprechtag aufgeschlüsselt nach Schularten

Individuelle und schulische Bedingungsfaktoren des Beratungsangebots von Lehrkräften

Bei der Frage danach, welche Faktoren das Beratungsangebot von Lehrkräften beeinflussen, können *individuelle Faktoren* auf Seiten der Lehrkraft, wie die Kompetenzüberzeugungen in Bezug auf Elternberatung, die professionellen Überzeugungen zur Lehrerrolle und der wahrgenommene Nutzen von Elternberatung, sowie *Faktoren auf schulischer Ebene*, wie die wahrgenommenen schulischen Rahmenbedingungen für die Elternarbeit und -beratung, das mittlere Kompetenzniveau der Schüler, die Komposition der Schülerschaft bezogen auf den sozioökonomischen Hintergrund und den Migrationshintergrund sowie die Schulart und die damit verbundenen pädagogischen Konzepte und Traditionen, unterschieden werden.

Zur Beantwortung der Frage, welche dieser Faktoren das Beratungsangebot der Lehrkräfte an Elternsprechtagen sowie zusätzlich zu Elternsprechtagen beeinflussen, wurden logistische Regressionsanalysen durchgeführt. Als abhängige Variablen wurden die *Anzahl der Gespräche am Elternsprechtag* sowie die *Anzahl der Gespräche zusätzlich zum Elternsprechtag* herangezogen. Die Ergebnisse der Analysen sind in Tabelle 3 dargestellt.

Prädiktoren	Beratungsangebot am Elternsprechtage			Beratungsangebot zusätzlich zum Elternsprechtage		
	b	(SE)	β	b	(SE)	β
<i>Individuelle Faktoren der Lehrkräfte</i>						
Kompetenzüberzeugung bezogen auf Elternberatung	-.002	.08	-.001	.06	.07	.02
Überzeugung zur Lehrerrolle	.12	.08	.05	.23	.06	.09
Überzeugung zum Nutzen von Elterngesprächen	.01	.06	.00	.09	.05	.05^(#)
<i>Schulische Faktoren</i>						
Wahrgenommene Rahmenbedingungen für Elternarbeit/-beratung an der Schule	.11	.07	.05	.58	.06	.26
Mittlere Lesekompetenz	.00	.00	.12	-.004	.001	-.22
Mittlerer sozioökonomischer Hintergrund	.01	.02	.06	.03	.01	.18
Anteil von Schülern mit Migrationshintergrund	.54	.34	.09	.71	.23	.11
Schularten mit Hauptschulbildungsgang	-.44	.19	-.44	-.24	.12	-.24
Gymnasium	-.35	.16	-.35	-.15	.10	-.15
R ²	.05			.16		

Anmerkung: Fettgedruckte Regressionskoeffizienten sind statistisch signifikant ($p < .05$); (#) = $p < .10$; Testung: zweiseitig.

Tab. 3: Regressionskoeffizienten der Zusammenhänge von individuellen und schulischen Faktoren mit dem Beratungsangebot von Deutschlehrkräften

Bezogen auf das Beratungsangebot am Elternsprechtage ergibt sich lediglich ein Einfluss der Schulart: Lehrkräfte führen weniger Gespräche am Elternsprechtage, wenn sie an einem Gymnasium oder an einer Schule mit Hauptschulbildungsgang unterrichten. Für die individuellen Faktoren auf Ebene der Lehrkraft sowie die weiteren schulischen Faktoren ergeben sich keine statistisch bedeutsamen Zusammenhänge mit dem Beratungsangebot am Elternsprechtage.

Für das Beratungsangebot zusätzlich zum Elternsprechtag ergeben sich ein statistisch signifikanter Zusammenhang mit dem individuellen Faktor *professionelle Überzeugung zur Lehrerrolle* sowie ein tendenzieller Zusammenhang mit dem wahrgenommenen Nutzen von Elterngesprächen ($p < .10$). Lehrkräfte, die die Beratung von Eltern stärker zu ihren Aufgaben zählen und den Nutzen von Elterngesprächen positiver beurteilen, führen häufiger Elterngespräche zusätzlich zum Elternsprechtag. Für die schulischen Faktoren ergeben sich statistisch signifikante Effekte bezogen auf die *wahrgenommenen Rahmenbedingungen für Elternarbeit und -beratung an der Schule*, die *mittlere Lesekompetenz* sowie den *mittleren sozioökonomischen Hintergrund* der Schüler und den *Anteil von Schülern mit Migrationshintergrund* an der Schule. Lehrkräfte, die an ihrer Schule mehr Ressourcen für die Elternarbeit wahrnehmen, an deren Schule die Schülerschaft eine niedrigere mittlere Lesekompetenz sowie einen höheren mittleren sozioökonomischen Hintergrund aufweist und häufiger einen Migrationshintergrund hat, führen häufiger Beratungsgespräche mit Eltern. Parallele Analysen unter Einbezug der Kompetenzbereiche Mathematik beziehungsweise Naturwissenschaften führen zu vergleichbaren Befunden.

4.4 Beratungswunsch der Eltern und Nutzung von Beratungs- und Informationsangeboten der Schule

Beratungswunsch der Eltern von Schülern in der 9. Jahrgangsstufe

Die Eltern der Schüler wurden danach gefragt, bei welchen Beratungsthemen Lehrkräfte wichtige Ansprechpersonen für sie sind. Für die Auswertungen wurden die Kategorien *stimmt überhaupt nicht* und *stimmt eher nicht* sowie die Kategorien *stimmt eher* und *stimmt ganz genau* zusammengefasst. Da die Angaben der Eltern über alle Schularten hinweg vergleichbar waren und sich nur geringfügig unterschieden, werden an dieser Stelle nur die Ergebnisse für die Gesamtstichprobe berichtet.

Die deskriptiven Analysen zeigen, dass ein Großteil der Schüler in Familien aufwächst, in denen Eltern Lehrkräfte insbesondere bei Fragen zum Leistungsstand des Kindes und zur Schulwahl (92.9%, $SE = 0.31$) sowie zur Versetzung des Kindes und zum Schulwechsel (92.4%, $SE = 0.39$) als wichtige Ansprechpersonen einstufen. Viele Schüler leben in Elternhäusern, in denen die Eltern die Bedeutung von Lehrkräften bei Fragen zum Umgang mit Lernschwierigkeiten des Kindes (87.4%, $SE = 0.48$), zur individuellen Förderung (78.0%, $SE = 0.69$) sowie zur Hilfestellung bei den Hausaufgaben (76.8%, $SE = 0.70$) als hoch beurteilen. Ein vergleichsweise geringerer Anteil der Schüler wächst in Familien auf, in denen die Eltern die Lehrkräfte als wichtige Ansprechpartner bei Fragen zum Umgang mit Verhaltensschwierigkeiten des Kindes und Erziehungsproblemen (69.7%, $SE = 0.67$), bei spezifischen Fragen von Eltern mit Migrationshintergrund (61.7%, $SE = 0.79$) sowie bei Fragen zu Sucht- und Drogenproblemen des Kindes (56.8%) ansehen. Dennoch werden diese Beratungsthemen in den Familien von mehr als der Hälfte der Schüler als wichtige Inhalte von Beratungsgesprächen mit Lehrkräften angesehen.

Nutzung von Beratungs- und Informationsangeboten

Um die Nutzung von Beratungs- und Informationsangeboten durch die Eltern zu erfassen, wurden diese gefragt, ob sie im letzten Schuljahr an einem Gesprächstermin oder einem Elternabend an der Schule ihres Kindes teilgenommen haben. Die Mehrheit der Schüler wächst in Familien auf, in denen die Eltern im letzten Schuljahr einen solchen Termin wahrgenommen haben (91.1%, $SE = 0.42$). Die differenzierte Betrachtung nach Schularten zeigt, dass der Anteil von Schülern, deren Eltern nicht an einem Gesprächstermin oder einem Elternabend teilgenommen haben, an Hauptschulen mit 12.9 Prozent ($SE = 1.09$) am höchsten ist, gefolgt von Schülern an Realschulen (10.0%, $SE = 0.88$), Schülern an Schulen mit mehreren Bildungsgängen (9.0%, $SE = 1.65$) und Schülern an Integrierten Gesamtschulen (8.4%, $SE = 1.31$). An Gymnasien ist der Anteil der Schüler, deren Eltern diese Form der Zusammenarbeit von Schule und Familie nicht wahrnehmen, am geringsten (6.4%, $SE = 0.51$).

Doch durch welche Faktoren wird die Teilnahme an Elterngesprächen und Elternabenden beeinflusst? Um diese Frage zu beantworten, wurden logistische Regressionsanalysen durchgeführt. Als abhängige Variable wurde die Teilnahme an Elterngesprächen bzw. Elternabenden herangezogen. Als unabhängige Variablen wurden zunächst individuelle Merkmale des Kindes (Geschlecht, Lesekompetenz, besuchte Schulart) aufgenommen (Modell I). In einem zweiten Modell wurden zusätzlich Struktur- und Prozessmerkmale der Familie – sozioökonomischer Hintergrund, Migrationshintergrund sowie elterliche Förderung des Kindes in der Grundschule und aktuelle Förderung durch die Eltern – aufgenommen. Die Ergebnisse der Auswertungen sind in Tabelle 4 dargestellt.

Werden nur die Schülermerkmale berücksichtigt (Modell I), ergeben sich für das *Geschlecht* ($p < .10$), die *Lesekompetenz* und die *Schulart* (tendenziell) statistisch signifikante Zusammenhänge mit der elterlichen Teilnahme an Beratungsgesprächen oder Elternabenden. Eltern nehmen häufiger an Elterngesprächen und -abenden teil, wenn ihr Kind männlichen Geschlechts ist, eine höhere Lesekompetenz aufweist und eine Schule mit Hauptschulbildungsgang besucht. Werden zusätzlich Struktur- und Prozessmerkmale der Familie berücksichtigt (Modell II), verschwinden diese Effekte jedoch und es ergeben sich keine statistisch signifikanten Zusammenhänge zwischen den Schülermerkmalen und der elterlichen Teilnahme an Beratungsgesprächen oder Elternabenden. Allerdings zeigen sich (tendenziell) statistisch signifikante Effekte für die Strukturmerkmale *sozioökonomischer Hintergrund* ($p < .10$) und *Migrationshintergrund* sowie für die Prozessmerkmale *Förderung in der Grundschule* und *aktuelle Förderung* durch die Eltern. Diese Befunde lassen sich wie folgt interpretieren: Eltern mit höherem sozioökonomischen Hintergrund nehmen häufiger an Beratungsgesprächen oder Elternabenden teil, Eltern aus Familien mit Migrationshintergrund hingegen seltener. Zudem nehmen Eltern, die ihr Kind während der Grundschulzeit intensiver gefördert haben und die ihr Kind aktuell unterstützen und fördern, häufiger an Elterngesprächen und -abenden teil. Das heißt: Sowohl der sozioökonomische und kulturelle Hintergrund als auch die Förderung des Kindes im Elternhaus sind für die Nutzung von Beratungs- und Informationsangeboten von Bedeutung.

Prädiktoren	Modell I			Modell II		
	b	(SE)	β	b	(SE)	β
Teilnahme an Beratungs- gesprächen und Eltern- abenden auf						
<i>Schülermerkmale</i>						
Geschlecht	.09	.05	.09(#)	.08	.06	.08
Lesekompetenz	.001	.00	.09	.00	.001	.03
Schulart mit Haupt- schulbildungsgang	.17	.08	.17	.11	.08	.11
Gymnasium	.01	.08	.01	.04	.08	.04
<i>Struktur- und Prozessmerkmale der Familie</i>						
Sozioökonomischer Hintergrund				.003	.002	.05(#)
Migrationshintergrund				-.47	.07	-.47
Förderung in der Grundschule				.15	.03	.12
Aktuelle Förderung				.12	.03	.10
R ²	.02			.09		

Anmerkung: Fettgedruckte Regressionskoeffizienten sind statistisch signifikant (p < .05); (#) = p < .10; Testung: zweiseitig.

Tab. 4: Regressionskoeffizienten der Zusammenhänge von Schülermerkmalen sowie Struktur- und Prozessvariablen der Familie mit der Teilnahme an Beratungsgesprächen bzw. Elternabenden

5. Diskussion und Fazit

Dieser Artikel zielte darauf ab, die Rahmenbedingungen für die Beratung von Eltern an Schulen im Sekundarbereich zu beschreiben und die Beratungsangebote von Lehrkräften sowie die Nutzung der Angebote durch die Eltern darzustellen. Den Hintergrund für die Analysen bildete ein Angebots-Nutzungs-Modell für die Elternberatung, welches auf der Grundlage theoretischer Überlegungen und empirischer Befunde entwickelt worden war.

Die Auswertungen zu den schulischen Rahmenbedingungen für die Elternberatung weisen darauf hin, dass diese einen hohen Stellenwert im Schulalltag einnimmt. Es gibt vielfältige Beratungsangebote für Eltern und an vielen Schulen wird Elternberatung zusätzlich zum Elternsprechtag durchgeführt (Forschungsfrage 1). Ein wichtiger Grund-

stein für die Zusammenarbeit von Schule und Eltern ist somit gelegt (Epstein et al., 2002; KMK, 2004). Positiv zu bewerten ist auch, dass nur ein sehr geringer Anteil der Schüler in der 9. Jahrgangsstufe in Familien aufwächst, in denen Eltern die Beratungs- und Informationsangebote der Schule nicht wahrnehmen.

Untersuchungen zur elterlichen Beteiligung an schulischen Bildungsprozessen geben Hinweise darauf, dass Eltern mit niedrigerem sozioökonomischen Status und Eltern mit Migrationshintergrund einen besonderen Beratungsbedarf aufweisen (z.B. Hill et al., 2004; Jeynes, 2007). Wenn Schulen auf diesen besonderen Bedarf reagieren, sollte sich dies in Zusammenhängen des Beratungsangebots und der Ressourcen für die Elternberatung der Schule mit der Zusammensetzung der Schüler- und Elternschaft niederschlagen (Forschungsfrage 2). Die Regressionsanalysen zeigen jedoch, dass sich nur vereinzelte Zusammenhänge der Beratungsangebote und entsprechender Ressourcen der Schule mit den Strukturmerkmalen der Schule (z.B. mittlere Lesekompetenz, mittlerer SES) ergeben. Die Schulen reagieren demnach bezogen auf Beratungsangebote und die Bereitstellung der Ressourcen noch nicht systematisch auf die Zusammensetzung ihrer Schüler- und Elternschaft. Vielmehr geben die Analysen Hinweise darauf, dass der Umfang des Beratungsangebots und die Verfügbarkeit struktureller Ressourcen für die Elternberatung insbesondere in einem Zusammenhang mit dem Stellenwert von Elternberatung an der Schule stehen.

In Hinblick auf die Beratungsangebote durch Lehrkräfte (Forschungsfrage 3) zeigt sich, dass fast alle Lehrkräfte sowohl am Elternsprechtag als auch zusätzlich zum Elternsprechtag Elternberatung anbieten. Nur sehr wenige Lehrkräfte (3.8%) unterrichten an Schulen, an denen es keinen Elternsprechtag gibt, und ein sehr geringer Anteil von Lehrkräften (0.5%) führt keine Elterngespräche zusätzlich zum Elternsprechtag. Die Regressionsanalysen für die Beratungsangebote am Elternsprechtag ergeben lediglich Effekte für die Schulart: An Schulen mit Hauptschulbildungsgang und an Gymnasien werden weniger Beratungsgespräche an einem Elternsprechtag geführt als an Realschulen. Die Auswertungen für die Beratungsangebote zusätzlich zum Elternsprechtag ergeben, dass sowohl individuelle Überzeugungen der Lehrkraft als auch die Ressourcen für Elternberatung an der Schule einen Einfluss auf das zusätzliche Beratungsangebot nehmen. Diese unterschiedlichen Befunde sind wenig verwunderlich, da der Elternsprechtag ein verpflichtendes Beratungs- und Informationsangebot darstellt. Individuelle Merkmale der Lehrkraft und strukturelle Merkmale der Schule werden erst relevant, wenn es sich um zusätzliche Angebote auf freiwilliger Basis handelt. Die Befunde stimmen mit den Ergebnissen einer Studie von Wild (2003) sowie dem Rahmenmodell zur Relevanz von Lehrerüberzeugungen für professionelles Handeln von Lehrkräften (Woolfolk Hoy, Davis & Pape, 2006) überein. Unerwartet ist allerdings, dass die Kompetenzüberzeugungen in Bezug auf Elternberatung unter Kontrolle des professionellen Rollenverständnisses sowie der Überzeugungen zum Nutzen von Elternsprachen keinen statistisch signifikanten Zusammenhang mit dem Beratungsangebot zusätzlich zum Elternsprechtag aufweisen. Methodisch könnte dies durch Multikollinearität bedingt sein. Die drei Überzeugungsmaße weisen statistisch signifikante Korrelationen zwischen $r = .36$ ($SE = .02$) und $r = .56$ ($SE = .02$) auf.

Für die Nutzung der Beratungsangebote der Schulen durch die Eltern (Forschungsfrage 4) zeigte sich Folgendes: Eltern mit niedrigerem sozioökonomischen Hintergrund sowie Eltern mit Migrationshintergrund nehmen entsprechende Angebote seltener wahr. Eltern, die ihr Kind während der Grundschulzeit und gegenwärtig stärker fördern, nehmen häufiger an Beratungsgesprächen mit Lehrkräften teil. Diese Befunde bestätigen die Vermutungen von Jeynes (2007), dass die Zusammenarbeit mit der Schule für Eltern mit niedrigem sozioökonomischen Hintergrund und für Eltern mit Migrationshintergrund mit besonderen Herausforderungen verbunden ist. Um welche Herausforderungen es sich hierbei spezifisch handelt, sollte im Rahmen weiterer Studien untersucht werden.

Ein Großteil der Zusammenhänge zwischen schulischen Rahmenbedingungen, Lehrervariablen und Merkmalen der Schüler- und Elternschaft, welche im Angebots-Nutzungs-Modell für die Elternberatung eingangs skizziert wurden, konnten durch die Analysen bestätigt werden. Die vorgestellten Auswertungen weisen allerdings einige Limitationen auf. Durch das querschnittliche Design können nur korrelative Aussagen getroffen werden. Eine Untersuchung kausaler Effekte ist nicht möglich. Dies ist mit maßgeblichen Einschränkungen für die Interpretation der Befunde verbunden. Zudem liegen nur Daten von Lehrkräften aus dem Fachkollegium Deutsch vor. Es können keine repräsentativen Aussagen über das gesamte Lehrerkollegium getroffen werden. Schließlich konnten keine Auswertungen zur Wirkung der elterlichen Beteiligung an Bildungsprozessen auf die Entwicklung von Verhalten, schulischen Leistungen und Aspirationen der Schüler erfolgen. Auch hierfür wäre ein Längsschnitt erforderlich gewesen. Inhaltlich ist anzumerken, dass in diesem Beitrag mit der Elternberatung nur ein spezifischer Aspekt der Beteiligung von Eltern an den schulischen Bildungsprozessen ihres Kindes und der Zusammenarbeit von Schule und Familie untersucht wurde. Eine Verallgemeinerung der Befunde auf weitere Aspekte der elterlichen Beteiligung ist nicht möglich.

Zusammenfassend weisen die Ergebnisse darauf hin, dass an Schulen in Deutschland günstige Rahmenbedingungen für die Beratung von Eltern vorliegen. Allerdings werden diese bisher nicht systematisch an die besonderen Beratungsbedarfe angepasst, welche sich aus der Zusammensetzung der Schüler- und Elternschaft ergeben. Eltern aus bildungsfernen Schichten und Eltern mit Migrationshintergrund werden seltener erreicht. Um die Beratungssituation zu verbessern, sollten sowohl Schulentwicklungsmaßnahmen als auch Aus- und Weiterbildungsprogramme für Lehrkräfte entwickelt, implementiert und evaluiert werden. Zudem sollten zukünftige Studien Unterschiede in Motivation, Gestaltung und Wirkung von elterlicher Beteiligung an Bildungsprozessen in Gruppen von Eltern unterschiedlicher sozioökonomischer und kultureller Herkunft untersuchen.

Literatur

- Aich, G. (2006). *Kompetente Lehrer: Ein Konzept zur Verbesserung der Konflikt- und Kommunikationsfähigkeit*. Hohengehren: Schneider.
- Bruder, S. (2011). *Lernberatung in der Schule. Ein zentraler Bereich professionellen Lehrerhandelns* (Dissertation). Darmstadt: Technische Universität Darmstadt, Institut für Psychologie, Fachbereich Humanwissenschaften.
- Bruder, S., Keller, S., Klug, J., & Schmitz, B. (2011). Ein Vergleich situativer Methoden zur Erfassung der Beratungskompetenz von Lehrkräften. *Unterrichtswissenschaft*, 39(2), 123-137.
- Catsambis, S. (1998). *Expanding knowledge of parental involvement in secondary education – Effects on high school academic success* (Johns Hopkins University, Report No. 27). Baltimore: CRESPAR (Center for Research on the Education of Students Placed at Risk). <http://www.csos.jhu.edu/crespar/techReports/Report27.pdf> [30.11.2012].
- Corno, L. (2000). Looking at homework differently. *Elementary School Journal*, 100, 529-548.
- Cox, D. D. (2005). Evidence based interventions using home school collaboration. *School Psychology Quarterly*, 20(4), 473-497.
- Ehmke, T., & Jude, N. (2010). Soziale Herkunft und Kompetenzerwerb. In E. Klieme, C. Artelt, J. Hartig, N. Jude, O. Köller, M. Prenzel, W. Schneider & P. Stanat (Hrsg.), *PISA 2009. Bilanz nach einem Jahrzehnt* (S. 231-254). Münster: Waxmann.
- Epstein, J. L., Sanders, M. G., Simon, B. S., Salinas, K. C., Jansorn, N. R., & van Voorhis, F. L. (2002). *School, family, and community partnerships: Your handbook for action* (2. Aufl.). Thousand Oaks: Corwin.
- Frey, A., Heinze, A., Mildner, D., Hochweber, J., & Asseburg, R. (2010). Mathematische Kompetenz von PISA 2003 bis PISA 2009. In E. Klieme, C. Artelt, J. Hartig, N. Jude, O. Köller, M. Prenzel, W. Schneider & P. Stanat (Hrsg.), *PISA 2009. Bilanz nach einem Jahrzehnt* (S. 153-176). Münster: Waxmann.
- Helmke, A. (2009). *Unterrichtsqualität und Lehrerprofessionalität. Diagnose, Evaluation und Verbesserung des Unterrichts*. Seelze: Klett/Kallmeyer.
- Hertel, S. (2009). *Beratungskompetenz von Lehrern. Kompetenzdiagnostik, Kompetenzförderung und Kompetenzmodellierung*. Münster: Waxmann.
- Hertel, S., & Djakovic, S.-K. (2010). *Die Beratungssituation an Grundschulen: Beratungsbedarf von Eltern und Beratungsangebote von Lehrpersonen* (Vortrag auf der 72. Tagung der Arbeitsgruppe für Empirische Pädagogische Forschung (AEPF) in Jena).
- Hertel, S., Hochweber, J., Steinert, B., & Klieme, E. (2010). Schulische Rahmenbedingungen und Lerngelegenheiten im Deutschunterricht. In E. Klieme, C. Artelt, J. Hartig, N. Jude, O. Köller, M. Prenzel, W. Schneider & P. Stanat (Hrsg.), *PISA 2009. Bilanz nach einem Jahrzehnt* (S. 113-151). Münster: Waxmann.
- Hertel, S., Jude, N., & Naumann, J. (2010). Leseförderung im Elternhaus. In E. Klieme, C. Artelt, J. Hartig, N. Jude, O. Köller, M. Prenzel, W. Schneider & P. Stanat (Hrsg.), *PISA 2009. Bilanz nach einem Jahrzehnt* (S. 255-275). Münster: Waxmann.
- Hertel, S., Steinert, B., Mildner, D., & Rauch, D. (in Vorb.). *PISA 2009: Dokumentation der Erhebungsinstrumente*. Münster: Waxmann.
- Hill, N. E., Castellino, D. R., Lansford, J. E., Nowlin, P., Dodge, K. A., Bates, J. E., & Pettit, G. S. (2004). Parental academic involvement as related to school behavior, achievement, and aspirations: Demographic variations across adolescence. *Child Development*, 75, 1491-1509.
- Hoover-Dempsey, K. V., Bassler, O. C., & Burow, R. (1995). Parents' report involvement in students' homework: Parameters of reported strategy and practice. *Elementary School Journal*, 95, 435-450.
- Hoover-Dempsey, K. V., & Sandler, H. M. (1997). Why do parents become involved in their children's education? *Review of Educational Research*, 67, 3-42.

- Jeynes, W. H. (2007). The relationship between parental involvement and urban secondary school student academic achievement: a meta-analysis. *Urban Education*, 42(1), 82-110.
- KMK = Kultusministerkonferenz (2004). *Standards für die Lehrerbildung: Bildungswissenschaften. Beschluss der Kultusministerkonferenz vom 16.12.2004.*
http://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2004/2004_12_16-Standards-Lehrerbildung.pdf [02.01.2012].
- Lengua, L., & McMahon, R. J. (2000). Parent Involvement in School Conceptualizing Multiple Dimensions and Their Relations with Family and Demographic Risk Factors. *Journal of School Psychology*, 38, 501-523.
- Muthén, L. K., & Muthén, B. O. (1998-2005). *Mplus User's Guide* (5. Aufl.). Los Angeles: Muthén & Muthén.
- Naumann, J., Artelt, C., Schneider, W., & Stanat, P. (2010). Lesekompetenz von PISA 2000 bis PISA 2009. In E. Klieme, C. Artelt, J. Hartig, N. Jude, O. Köller, M. Prenzel, W. Schneider & P. Stanat (Hrsg.), *PISA 2009. Bilanz nach einem Jahrzehnt* (S. 23-71). Münster: Waxmann.
- OECD (2011). *PISA 2009. Technical Report*. Paris: OECD.
- Rönnebeck, S., Schöps, K., Prenzel, M., Mildner, D., & Hochweber, J. (2010). Naturwissenschaftliche Kompetenz von PISA 2006 bis PISA 2009. In E. Klieme, C. Artelt, J. Hartig, N. Jude, O. Köller, M. Prenzel, W. Schneider & P. Stanat (Hrsg.), *PISA 2009. Bilanz nach einem Jahrzehnt* (S. 177-198). Münster: Waxmann.
- Schmitz, B., Bruder, R., & Hertel, S. (2011). *Modellierung der Beratungskompetenz von Lehrer/innen unter besonderer Berücksichtigung der Diagnostischen Kompetenz* (Fortsetzungsantrag auf Gewährung einer Sachbeihilfe im Rahmen des DFG-Schwerpunktprogramms: Kompetenzmodelle zur Erfassung individueller Lernergebnisse und zur Bilanzierung von Bildungsprozessen).
- Stanat, P., Rauch, D., & Segeritz, M. (2010). Schülerinnen und Schüler mit Migrationshintergrund. In E. Klieme, C. Artelt, J. Hartig, N. Jude, O. Köller, M. Prenzel, W. Schneider & P. Stanat (Hrsg.), *PISA 2009. Bilanz nach einem Jahrzehnt* (S. 200-230). Münster: Waxmann.
- Wild, E. (2003). Lernen lernen. Wege einer Förderung der Bereitschaft und Fähigkeit zu selbst-reguliertem Lernen. *Unterrichtswissenschaften*, 31, 2-5.
- Woolfolk Hoy, A., Davis, H., & Pape, S. J. (2006). Teacher knowledge and beliefs. In P. A. Alexander & P. H. Winne (Hrsg.), *Handbook of Educational Psychology* (S. 715-737). Mahwah: Lawrence Erlbaum.

Anschrift der Autorinnen

JunProf. Dr. Silke Hertel, Deutsches Institut für Internationale Pädagogische Forschung (DIPF), Schloßstraße 29, 60486 Frankfurt am Main, Deutschland
 E-Mail: hertel@dipf.de

Dr. Simone Bruder, Kinderkliniken Prinzessin Margaret, Dieburger Str. 31, 64287 Darmstadt, Deutschland

Dr. Nina Jude, Deutsches Institut für Internationale Pädagogische Forschung (DIPF), Schloßstr. 29, 60486 Frankfurt am Main, Deutschland
 E-Mail: jude@dipf.de

Dr. Brigitte Steinert, Deutsches Institut für Internationale Pädagogische Forschung (DIPF), Schloßstraße 29, 60486 Frankfurt am Main, Deutschland
 E-Mail: steinert@dipf.de

Timo Ehmke

Soziale Disparitäten im Lesen und in Mathematik innerhalb von Schulklassen

1. Einleitung

Die enge Kopplung zwischen den erreichten Kompetenzen von Jugendlichen in Deutschland und den Merkmalen ihrer sozialen Herkunft ist insbesondere seit PISA 2000 wieder Gegenstand empirischer Untersuchungen und bildungspolitischer Diskussionen. Die regelmäßige Auswertung der Daten der PISA-Studien zeigt, dass sich die Situation von Jugendlichen aus den unteren sozialen Schichten von PISA 2000 nach PISA 2009 insgesamt verbessert hat. Dennoch erreichen Jugendliche mit einem niedrigen sozioökonomischen Status durchschnittlich immer noch deutlich geringere Kompetenzen als Gleichaltrige, deren Eltern einen höheren sozioökonomischen Status aufweisen (OECD, 2010; Ehmke & Jude, 2010). Neben der regelmäßigen Beschreibung und Quantifizierung von sozialen Disparitäten des Kompetenzerwerbs ist aber von wesentlichem Interesse, welche bildungspolitischen und schulischen Maßnahmen dazu beitragen, um diese abzuschwächen oder gar nicht erst entstehen zu lassen. Maaz, Baumert & Trautwein (2009) haben hier vier Erklärungsansätze unterschieden, die für die Genese sozialer Ungleichheit im institutionellen Kontext der Schule relevant sind. Der erste Erklärungsansatz bezieht sich auf das Entscheidungsverhalten bei Bildungsübergängen und verweist darauf, dass soziale Ungleichheiten in der Bildungsbeteiligung zu einem Teil durch sozialschichtabhängige Entscheidungen der Eltern und durch ein sozial selektives Beratungs- und Empfehlungsverhalten von Lehrkräften entstehen (Boudon, 1974; Ditton, Krüsken & Schauenberg, 2005; Maaz & Nagy, 2009). Zweitens können soziale Disparitäten des Kompetenzerwerbs auch außerhalb des institutionellen Bildungssystems verstärkt werden. Ein Beispiel hierfür ist der sogenannte Sommerferieneffekt, der sozial differenzielle Lernentwicklungen während der Sommerferien beschreibt. Demnach profitieren Kinder aus anregungsreicheren Herkunftsmilieus über die Sommerferien stärker in ihrer Lernentwicklung als Kinder aus weniger anregungsreichen Herkunftsfamilien (Cooper, Nye, Charlton, Lindsay & Greathouse, 1996; Becker, Stanat, Baumert & Lehmann, 2008). Als Drittes ist die Entstehung von sozialen Disparitäten *zwischen* Bildungsinstitutionen zu nennen. Eine Ursache für solche institutionellen Effekte liegt in der sozialschichtabhängigen Zusammensetzung der Schülerinnen und Schüler in unterschiedlichen Schulformen. Mit solchen sozial unterschiedlichen Klassenkompositionen sind differenzielle Lern- und Entwicklungsmilieus verbunden, die zu einer unterschiedlichen Leistungsentwicklung zwischen den Schulformen führen können (Baumert, Stanat & Watermann, 2006; Maaz, Trautwein, Lüdtke & Baumert, 2008; Hattie, 2002; Opdenakker & van Damme, 2001; Thrupp, Lauder & Robinson, 2002). Letztlich können soziale Disparitäten des Kompetenzerwerbs aber auch *innerhalb* von Bildungsin-

stitutionen entstehen, also innerhalb von Schulen oder von Schulklassen. Aus schulpädagogischer und bildungspolitischer Perspektive ist insbesondere dieser letzte Bereich relevant. Denn im Gegensatz zu den anderen Entstehungsmustern könnten hier prinzipiell pädagogische Interventionen ansetzen oder Routinen innerhalb von Schulen verändert werden, um gewünschte Entkopplungseffekte zu erzielen. In der Forschung wurde dieser Ursachenbereich noch relativ wenig untersucht (Maaz, Baumert & Trautwein, 2009). Auf bislang vorliegende theoretische Annahmen und Befunde wird im Folgenden eingegangen.

1.1 Entstehung von sozialen Disparitäten innerhalb von Schulklassen

Wie können soziale Disparitäten beim Kompetenzerwerb erklärt werden, die innerhalb von Schulklassen entstehen? Eine erste Annahme bezieht sich auf eine Wechselwirkung zwischen Herkunftsmerkmalen der Schülerinnen und Schüler und dem Angebot bzw. der effektiven Nutzung von schulischen Lerngelegenheiten. Schulische Ressourcen und Lerngelegenheiten werden möglicherweise von Schülerinnen und Schülern aus verschiedenen sozialen Schichten unterschiedlich effektiv genutzt (vgl. Maaz, Baumert & Trautwein, 2009). Die theoretischen Arbeiten von Bourdieu & Passeron (1971) erklären diese Wechselwirkung durch eine bessere Passung zwischen den geforderten schulischen Verhaltensnormen und Sprachcodes einerseits und dem sozialen und kulturellen Habitus von Schülerinnen und Schülern aus mittleren und oberen Sozialschichten andererseits. Kinder und Jugendliche aus unteren Sozialschichten weisen nach diesem Ansatz einen auf schulische Anforderungen weniger passenden kulturellen Habitus auf, was zu einer weniger effektiven Nutzung der schulischen Lerngelegenheiten führt (vgl. Bourdieu, 1982).

Ein zweiter Erklärungsansatz besteht nach Maaz, Baumert & Trautwein (2009, S. 80) darin, „dass eine sozial (und ethnisch) selektive Erwartungs-, Wertschätzungs- und Belohnungsstruktur in Bildungsinstitutionen für über die Schullaufbahn wachsende soziale Disparitäten verantwortlich ist (vgl. Lehmann, Peek & Gänsfuß, 1997)“. So konnte beispielsweise anhand der Daten aus PISA 2003 gezeigt werden, dass Eltern mit höherem Bildungsabschluss auch eine höhere fachspezifische Wertschätzung aufweisen, eine ausgeprägtere Lernunterstützung bieten und anspruchsvollere Bildungsaspirationen verfolgen als Eltern, die eher niedrigere Bildungsabschlüsse aufweisen (Ehmke & Siegle, 2008; Ehmke, 2008).

Über diese beiden Erklärungsansätze hinaus kann angenommen werden, dass die Wechselwirkung zwischen der sozialen Herkunft der Schülerinnen und Schüler und dem schulischen Lernerfolg innerhalb einer Klasse sowohl von strukturellen Merkmalen der Klassenzusammensetzung als auch von prozessorientierten Unterrichts- oder Schulmerkmalen beeinflusst wird (Steuer, Berner & Dresel, 2010). Zu strukturellen Merkmalen zählen alle Aspekte, die die Zusammensetzung der Schulklassen charakterisieren, wie etwa der mittlere sozioökonomische Status, der Anteil von Schülerinnen und Schülern mit nicht-deutscher Familiensprache oder die Leistungsheterogenität inner-

halb einer Klasse. Unter prozessorientierten Merkmalen des Unterrichts oder der Schule können Förder- oder Unterrichtsaspekte verstanden werden, die auf eine eher kompensatorische Wirkung abzielen. Dazu zählen etwa das Gewähren von inner- und außerunterrichtlichen Fördermaßnahmen oder das Anbieten von Hausaufgabenbetreuungen und anderen lernförderlichen Ganztagsangeboten.

1.2 Empirische Befunde zu sozialen Disparitäten innerhalb von Schulklassen

Wie hoch sind in Deutschland die sozialen Disparitäten des Kompetenzerwerbs *innerhalb* von Schulklassen im Vergleich zu den sozialen Unterschieden *zwischen* den Schulen ausgeprägt? Repräsentative Befunde für die Fünfzehnjährigen in Deutschland liegen aus PISA 2009 vor. Das Verhältnis dieser beiden Disparitätsmaße kann anhand der Steigungen der sozialen Gradienten innerhalb und zwischen Schulen abgeschätzt werden. Die Befunde (OECD, 2010, S. 115) ergaben hier, dass bei den Fünfzehnjährigen in Deutschland die sozialen Gradienten zwischen den Schulen bedeutsam steiler ausfallen als im Durchschnitt der OECD-Staaten. Mehr als drei Viertel der Unterschiede in der mittleren Lesekompetenz zwischen den Schulen lassen sich durch Unterschiede in dem durchschnittlichen sozialen Hintergrund der Schülerinnen und Schüler auf Schulebene erklären. Demnach kann man in Deutschland leistungsstarke Schulen, deren Schülerinnen und Schüler einen hohen mittleren sozioökonomischen Status aufweisen, relativ deutlich von eher leistungsschwachen Schulen abgrenzen, deren Schülerinnen und Schüler überwiegend aus weniger privilegierten Herkunftsfamilien stammen. Entsprechend unterscheiden sich auch die Steigungen der sozialen Gradienten der Lesekompetenz innerhalb und zwischen den Schulen. Während für Deutschland die Steigung des sozialen Gradienten der Lesekompetenz innerhalb der Schulen bei 10 Punkten liegt, fällt die Steigung des sozialen Gradienten zwischen den Schulen mit 122 Punkten um ein Vielfaches höher aus (OECD, 2010, S. 189-190).

Die hohe Kopplung zwischen sozialer Herkunft und Lesekompetenz kann in Deutschland somit vor allem über den herkunftsspezifischen Besuch bestimmter Schulen bzw. bestimmter Schulformen erklärt werden. Dennoch lässt sich in PISA 2009 auch innerhalb der Schulen ein nachweisbarer Herkunftseffekt feststellen. Welche Bedingungsfaktoren sind nun für eine hohe oder geringe Kopplung innerhalb von Schulen bzw. Klassen ausschlaggebend? Ma & Klinger (2000) und Ma (2000) konnten hierzu an einer kanadischen Stichprobe von Sechstklässlern zeigen, dass die innerschulische Steigung der sozialen Gradienten für die Mathematikleistung umso stärker ausfiel, je mehr Schülerinnen und Schüler eine Schule besuchten und je stärker die schulbezogene Unterstützung der Eltern ausfiel. Dieser Zusammenhang ließ sich nur für die Mathematikleistung, aber nicht für die Domänen Naturwissenschaften, Lesen oder Schreiben nachweisen. Auch waren das disziplinäre Klima an den Schulen und der schulische Leistungsdruck keine bedeutsamen Prädiktoren für die Höhe der innerschulischen sozialen Disparitäten.

In einer deutschsprachigen Studie zeigten Steuer, Berner & Dresel (2010), dass sich keine Effekte von strukturellen Merkmalen, gemessen am Anteil von Schülerinnen und Schülern mit nicht-deutscher Familiensprache sowie am mittleren sozioökonomischen Status, auf das klassenspezifische Ausmaß von Leistungsdisparitäten belegen lassen. Hingegen fand sich in einer Teilstudie ein Effekt der individuellen Bezugsnormorientierung der Lehrkraft auf den klassenspezifischen Zusammenhang zwischen dem sozioökonomischen Status und der Schulleistung.

Inwieweit innerschulische soziale Disparitäten konsistent zwischen fachlichen Domänen sind, wurde von Ma (2000) untersucht. Die Höhe der sozialen Gradienten innerhalb von Schulen korrelierte signifikant zwischen den Domänen Mathematik, Naturwissenschaften, Lesen und Schreiben ($0.54 < r < 0.82$). Die engsten Zusammenhänge ergaben sich für die innerschulischen sozialen Gradienten im Lesen und im Schreiben ($r = .82$) und in Mathematik und in den Naturwissenschaften ($r = .70$). In einer weiteren Studie konnte ein enger Zusammenhang zwischen den innerschulischen sozialen Gradienten in Mathematik und Naturwissenschaften ebenfalls bestätigt werden (Ma, 2001). Insgesamt liegen somit erste Befunde zur Höhe von klassen- und schulspezifischen sozialen Disparitäten in schulischen Leistungen vor. Inwieweit diese zusätzlich durch kompensatorische Fördermaßnahmen im Unterricht oder in Schulen abgeschwächt werden können, ist bislang noch weitgehend ungeklärt (vgl. auch Ditton, 2010).

2. Fragestellungen

Ziel dieser Studie ist es, die klassenspezifischen sozialen Disparitäten im Lesen und in der Mathematik an einer repräsentativen Stichprobe von Neuntklässlern aus Deutschland zu untersuchen und folgenden Fragestellungen nachzugehen:

1. Wie hoch fallen die klassenspezifischen sozialen Disparitäten im Lesen und in der Mathematik für Schülerinnen und Schüler in der neunten Klassenstufe aus?
2. Inwieweit fallen die klassenspezifischen sozialen Disparitäten in den Domänen Mathematik und Lesen konsistent aus?
3. Inwieweit lassen sich klassenspezifische soziale Disparitäten durch strukturelle Klassenmerkmale und unterrichtsbezogene und außerschulische Fördermaßnahmen erklären?

Zu 1.: Anhand der repräsentativen Stichprobe der Neuntklässler aus PISA 2009 soll in einer ersten deskriptiven Analyse insgesamt und getrennt nach Schulformen ausgewertet werden, wie hoch der klassenspezifische Zusammenhang zwischen den erreichten Kompetenzen im Lesen oder in der Mathematik und der sozialen Herkunft der Jugendlichen ausfällt. Es ist anzunehmen, dass sich statistisch bedeutsame klassenspezifische Disparitäten nachweisen lassen. Es kann weiter vermutet werden, dass aufgrund der sozial selektiven Bildungsbeteiligung die klassenspezifischen Disparitäten zwischen den Schulformen variieren. Ferner soll geprüft werden, ob die klassenspezifischen Dispari-

täten in der mathematischen Kompetenz höher oder geringer ausfallen als in der Lesekompetenz. Für beide Richtungen lassen sich theoretische Argumente finden. Für höhere klasseninterne Disparitäten im Lesen als in der Mathematik lässt sich theoretisch argumentieren, dass die Lesekompetenz zu einem hohen Anteil auch außerschulisch erworben wird und somit stärker von der familiären kulturellen Praxis abhängig ist. Für höhere klasseninterne Disparitäten in der Mathematik als im Lesen spricht dagegen, dass gerade im mathematischen Bereich die häusliche Unterstützung etwa bei den Hausaufgaben relevant ist und dass Unterschiede in der sozialen Herkunft damit stärker zum Tragen kommen könnten. Auch zielen unterrichtliche und schulische Förderungen vor allem auf die sprachlichen Fähigkeiten ab und liegen weniger im mathematisch-naturwissenschaftlichen Bereich.

Zu 2.: Um zu untersuchen, inwieweit die Höhe der klassenspezifischen sozialen Disparitäten konsistent über die Domänen Mathematik und Lesen ist, wird der Zusammenhang zwischen den Gradientensteigungen im Lesen und in der Mathematik auf der Klassenebene analysiert. Sollte dieser Zusammenhang gering ausfallen, könnte dies ein Hinweis darauf sein, dass klassenspezifische soziale Disparitäten vor allem durch fachspezifische oder lehrkraftabhängige Unterrichtsmerkmale oder eine domänenspezifische Förderung bedingt sind. Ein enger Zusammenhang würde hingegen stärker die Annahme stützen, dass Kinder und Jugendliche aus höheren Sozialschichten allgemein eine effektivere Nutzung der schulischen Lerngelegenheiten gelingt.

Zu 3.: In dem dritten Auswertungsschritt werden mögliche Bedingungsfaktoren für klassenspezifische soziale Disparitäten anhand von Slope-As-Outcome-Modellen analysiert (Raudenbush & Bryk, 2002). Relevante Prädiktoren können sich zum einen auf strukturelle Klassenmerkmale beziehen, die die Zusammensetzung der Schulklassen anhand von Kontextvariablen wie dem Migrationsanteil charakterisieren, oder zum anderen eher unterrichtsbezogene oder außerunterrichtliche Förderangebote umfassen. Eine höhere Vorhersagekraft ist eher für die unterrichtsbezogenen oder außerunterrichtlichen Förderangebote zu erwarten. So fallen klassenspezifische Leistungsdisparitäten möglicherweise geringer aus, wenn sich der Unterricht durch einen binnendifferenzierten Umgang mit leistungsschwachen Schülerinnen und Schülern auszeichnet (vgl. Bos, Dohe, Kowoll, Platz & Schuster, 2010; Bos & Scharenberg, 2010). Auch ist zu erwarten, dass unterrichtliche Kompensationsmaßnahmen und schulische Nachmittagsangebote wie Hausaufgabenbetreuung zu einer Reduktion von klasseninternen Leistungsunterschieden beitragen.

3. Datengrundlage und Methodik

3.1 Stichprobe

Die Studie basiert auf den Daten der nationalen Erweiterung der PISA-Stichprobe von Schülerinnen und Schülern der 9. Klassenstufe in Deutschland (Klieme et al., 2010). Neben den Kompetenzmessungen bei den Jugendlichen im Lesen, in Mathematik und in den Naturwissenschaften wurden auch Eltern, Lehrkräfte und Schulleitungen so-

wie die Schülerinnen und Schüler befragt, um die Rahmenbedingungen des schulischen Lernens zu erfassen (Jude & Klieme, 2010). Insgesamt haben in Deutschland $N = 9461$ Schülerinnen und Schüler aus 393 Schulklassen der neunten Jahrgangsstufe an der Studie teilgenommen. Von den Analysen ausgeschlossen wurden vier Schulklassen, die weniger als 10 Schülerinnen und Schüler umfassten. Eine ausreichende Stichprobengröße für die Schätzungen der klassenspezifischen Disparitäten wurde somit sichergestellt. Die finale Klassenstichprobe umfasst 77 Schulklassen von Hauptschulen, 30 Klassen von Schulen mit mehreren Bildungsgängen, 111 Klassen von Realschulen, 36 Klassen von Integrierten Gesamtschulen und 135 Gymnasialklassen. Ergänzende Informationen zum Erhebungsdesign und zur Stichprobenziehung von PISA 2009 in Deutschland geben Jude & Klieme (2010).

3.2 Auswertungsmethoden

Um zu überprüfen, inwieweit sich Schulklassen unterschiedlicher Schulformen hinsichtlich des Zusammenhangs von erreichten Kompetenzen und sozialer Herkunft der Jugendlichen unterscheiden, werden in einem ersten Schritt Regressionsanalysen auf der Ebene von Schulklassen durchgeführt. Im nächsten Auswertungsschritt wird dann der Zusammenhang zwischen den klassenspezifischen Disparitäten in Mathematik und im Lesen anhand eines Streudiagramms grafisch veranschaulicht und anhand des Korrelationsmaßes quantifiziert. Im dritten Schritt wird für die Analyse von Bedingungsfaktoren für die klassenspezifischen Disparitäten ein mehrebenenanalytisches Vorgehen gewählt. Hierbei wird ein Vorhersagemodell auf zwei Ebenen spezifiziert, das neben Individualmerkmalen auch Klassenmerkmale einbezieht. Im Folgenden soll anhand eines vereinfachten Beispiels kurz dargestellt werden, wie die Vorhersage der Gradientensteigungen durch Klassenmerkmale im Rahmen des mehrebenenanalytischen Ansatzes modelliert werden kann. Diese Darstellung basiert auf einer Beschreibung von Kontextanalysen in einem Beitrag von Stanat (2006; vgl. auch Lüdtke, Robitzsch & Köller, 2002).

Im folgenden Modell wird auf der Individualebene angenommen, dass sich die Leistung Y eines Schülers i in der Schulklasse j aus dem Achsenabschnittsparameter β_{0j} der Schulklasse j , dem Einfluss β_{1j} des sozioökonomischen Status des Schülers i in der Schulklasse j , dem Einfluss β_{2j} des Vorwissens des Schülers i in der Schulklasse j und einem Residualterm e_{ij} zusammensetzt:

$$(1) Y_{ij} = \beta_{0j} + \beta_{1j} (\text{soz. Status})_{ij} + \beta_{2j} (\text{Vorwissen})_{ij} + e_{ij}$$

In dieser Gleichung stellt der Achsenabschnittsparameter β_{0j} die mittlere Leistung in der Schulklasse j dar, korrigiert um die Effekte des sozioökonomischen Status und des Vorwissens auf der Individualebene. Der Achsenabschnittsparameter β_{0j} wird als Variable behandelt, die zwischen Schulklassen variieren kann:

$$(2) \beta_{0j} = \gamma_{00} + u_{0j}$$

Auf der Klassenebene wird der Steigungsparameter β_{1j} , der den sozialen Gradienten innerhalb einer Schulklasse j beschreibt, als abhängige Variable behandelt, die zwischen Schulklassen variieren kann. Soll nun beispielsweise untersucht werden, ob die Steigung des sozialen Gradienten vom Angebot einer Fördermaßnahme (Förderangebot) abhängt, so lautet die entsprechende Gleichung:

$$(3) \quad \beta_{1j} = \gamma_{10} + \gamma_{11} (\text{Förderangebot})_j + u_{1j}$$

Der Steigungsparameter β_{1j} wird also vorhergesagt durch den Koeffizienten γ_{10} , der die mittlere Steigung der sozialen Gradienten über alle Schulklassen hinweg repräsentiert, durch den Einfluss γ_{11} des Förderangebots in den Schulklassen und durch einen klassenspezifischen Residualterm u_{1j} . Zugleich muss das Klassenmerkmal „Förderangebot“ jetzt aber auch bei der Schätzung des Achsenabschnittsparameters β_{0j} in Gleichung (2) mit berücksichtigt werden.

Für den Regressionskoeffizienten des Vorwissens wird in diesem Fall angenommen, dass dieser nicht zwischen den Schulklassen variiert:

$$(4) \quad \beta_{2j} = \gamma_{20}$$

Setzt man die Gleichungen (2) bis (4) in die Gleichung (1) ein, so ergibt sich:

$$(5) \quad Y_{ij} = \gamma_{00} + \gamma_{01} (\text{Förderangebot})_j + \gamma_{10} (\text{soz. Status})_{ij} + \gamma_{11} (\text{Förderangebot})_j (\text{soz. Status})_{ij} + \gamma_{20} (\text{Vorwissen})_{ij} + u_{0j} + u_{1j} + e_{ij}$$

Bei den Modellen, die in den folgenden Analysen geschätzt werden, handelt es sich um Erweiterungen dieser Gleichung, in die zusätzliche Variablen auf der Klassenebene eingehen. Die Schätzung der Modellparameter erfolgt mit dem Programm HLM 6.06 (Raudenbush, Bryk & Congdon, 2004). Für alle statistischen Analysen wird ein Signifikanzniveau von $p < 0.05$ festgelegt.

Für die Auswertungen werden vollständige Datensätze verwendet. Fehlende Werte wurden nachträglich mit der Software *Norm* 2.03 (Schafer, 2000; Schafer & Graham, 2002) geschätzt.

3.3 Variablen

Die mathematische Kompetenz und die Lesekompetenz wurden durch die Testinstrumente aus PISA 2009 erfasst, die auf den PISA-Rahmenkonzeptionen zu Mathematical Literacy und Reading Literacy beruhen (vgl. OECD, 2009). Die Testdaten wurden anhand eines eindimensionalen Raschmodells skaliert und liegen in der Testmetrik des internationalen OECD-Vergleichs vor (OECD, 2012).

Der sozioökonomische Status der Familien wird anhand des von Ganzeboom, de Graaf, Treiman und de Leeuw (1992) entwickelten *International Socio-Economic Index* (ISEI) erfasst. Der Index basiert auf der Angabe des zuletzt ausgeübten Berufs der erwachsenen Bezugspersonen. Die Berufsangaben der Eltern wurde zuvor anhand der *International Standard Classification of Occupation* von 1988 (ISCO-88, International Labor Office, 1990) klassifiziert. In allen Analysen wird der höchste ISEI-Wert in der Familie (HISEI = Highest ISEI) genutzt (vgl. Ehmke & Siegle, 2005).

Als Indikator für das Vorwissen der Schülerinnen und Schüler wird ein Maß für die kognitiven Grundfähigkeiten verwendet (Heller & Perleth, 2000), die mit zwei Subskalen des Kognitiven Fähigkeitstests (KFT) erfasst wurden. Der Test umfasst insgesamt 25 Items und weist eine Reliabilität von Cronbach's $\alpha = 0.87$ auf.

Der Migrationshintergrund der Schülerinnen und Schüler wird am Geburtsland der Eltern festgemacht. Dabei werden zwei Gruppen unterschieden: (1.) Familien, in denen beide Eltern oder wenigstens ein Elternteil in Deutschland geboren ist bzw. sind, und (2.) Familien, in denen beide Eltern im Ausland geboren sind.

Auf der Klassenebene wird in den Mehrebenenmodellen zwischen strukturellen Klassenmerkmalen und unterrichtsbezogenen und außerschulischen Fördermaßnahmen unterschieden. Als strukturelle Merkmale werden die mittleren kognitiven Fähigkeiten der Schülerinnen und Schüler, das durchschnittliche Alter der Jugendlichen und der prozentuale Anteil von Jugendlichen mit Migrationshintergrund berücksichtigt. Um die Heterogenität der Lernvoraussetzungen in den Schulklassen zu quantifizieren, werden zudem die klassenspezifischen Standardabweichungen in den kognitiven Fähigkeiten und im Alter der Schülerinnen und Schüler einbezogen. Als Indikatoren für unterrichtsbezogene und außerschulische Fördermaßnahmen werden auf der Klassenebene fünf Indikatoren betrachtet, die anhand des Schülerfragebogens erhoben und anschließend auf Klassenebene aggregiert wurden:

- (1) Zusatzunterricht für leistungsstarke Schülerinnen und Schüler im Fach Deutsch oder Mathematik (2 Items, Beispiel: „Besuchst du derzeit Zusatz- oder Förderunterricht außerhalb der normalen Schulstunden? – Zusatzunterricht für leistungsstarke Schülerinnen/Schüler im Fach Deutsch“).
- (2) Zusatzunterricht für leistungsschwache Schülerinnen und Schüler im Fach Deutsch oder Mathematik (2 Items, Beispiel: „Besuchst du derzeit Zusatz- oder Förderunterricht außerhalb der normalen Schulstunden? – Förderunterricht/Nachhilfeunterricht für Schülerinnen/Schüler im Fach Deutsch“).
- (3) Privater Einzelunterricht (Einzelitem, Beispiel: „Besuchst du derzeit Zusatz- oder Förderunterricht außerhalb der normalen Schulstunden? – Privater Einzelunterricht/Private Einzelnachhilfe“).
- (4) Schulische Hausaufgabenbetreuung (Einzelitem, Beispiel: „Besuchst du derzeit Zusatz- oder Förderunterricht außerhalb der normalen Schulstunden? – Hausaufgabenbetreuung/Hausaufgabenhilfe“).

- (5) Teilnahme am schulischen Ganztagsangebot (Einzelitem, Beispiel: „Zunächst möchten wir wissen, ob du die Ganztagsangebote an deiner Schule besuchst. – Ja, ich besuche die Ganztagsangebote.“).

4. Ergebnisse

4.1 *Klassenspezifische soziale Disparitäten im Lesen und in der Mathematik*

Wie hoch fällt der Zusammenhang zwischen der erreichten Mathematik- und Lesekompetenz und der sozialen Herkunft von Jugendlichen innerhalb von Schulklassen aus? Die Tabellen 1 und 2 stellen für die beiden Kompetenzdomänen Lesen und Mathematik deskriptive Kennwerte zu den sozialen Gradienten innerhalb von Schulklassen getrennt nach Schulformen dar. Zusätzlich sind die Kennwerte für die Gesamtstichprobe angegeben. Als Maß für die soziale Herkunft wird der höchste sozioökonomische Status der Eltern (HISEI) in der Analyse verwendet.

Für die Gesamtheit der neunten Klassen ergibt sich eine durchschnittliche Steigung der klassenspezifischen sozialen Gradienten in der Lesekompetenz von 7 Punkten, in der mathematischen Kompetenz von 8 Punkten. Erwartungsgemäß fällt die durchschnittliche Steigung der sozialen Gradienten innerhalb der Schulklassen damit deutlich geringer aus als für die Gesamtstichprobe, wenn die Klassenstruktur nicht berücksichtigt wird (vgl. Ehmke & Jude, 2010). Allerdings ist eine erhebliche Unterschiedlichkeit zwischen den Klassen zu beobachten. In der Domäne Lesen variieren die Steigungen der Gradienten innerhalb der Schulklassen zwischen –53 und +74 Punkten, in der Mathematik zwischen –33 und +88 Punkten. Es gibt demnach Schulklassen mit deutlich ausgeprägten Herkunftseffekten. Es finden sich aber auch Schulklassen, in denen ein höherer sozioökonomischer Status mit geringeren Kompetenzwerten bei den Schülerinnen und Schülern einhergeht.

Das Ergebnismuster der gesamten Klassenstichprobe zeigt sich auch für die einzelnen Schulformen, wenngleich es spezifische Unterschiede gibt. Die höchsten Steigungen in den klassenspezifischen sozialen Gradienten finden sich in den Integrierten Gesamtschulen (Gradientensteigung im Lesen: 21 Punkte, Mathematik: 18 Punkte) und in den Schulen mit mehreren Bildungsgängen (Lesen: 14 Punkte, Mathematik: 11 Punkte). Im Gegensatz dazu sind die klassenspezifischen sozialen Gradienten in den Hauptschulen, Realschulen und Gymnasien im Durchschnitt niedriger ausgeprägt. In allen Schulformen lassen sich aber bedeutsame Schwankungen in den Gradientensteigungen feststellen (SD-Werte in den Tabellen 1 und 2). Auch gibt es Klassen, in denen negative Gradientensteigungen zu beobachten sind. In diesen Klassen erreichen Jugendliche aus unteren Sozialschichten höhere Kompetenzwerte als Kinder, deren Eltern einen höheren sozioökonomischen Status aufweisen.

	MW	SD	Min	Max
Hauptschule	4.7	20.6	-52.5	56.2
Schulen mit mehreren Bildungsgängen	13.6	22.3	-26.5	74.0
Realschule	6.3	13.7	-21.2	59.7
Integrierte Gesamtschulen	21.3	14.1	-2.7	53.8
Gymnasium	4.5	12.9	-30.9	43.7
Gesamt	7.3	16.6	-52.5	74.0

Tab. 1: Soziale Gradienten der Lesekompetenz innerhalb von Schulen differenziert nach Schulformen (unstandardisierte Regressionskoeffizienten)

	MW	SD	Min	Max
Hauptschule	5.8	18.1	-33.4	55.0
Schulen mit mehreren Bildungsgängen	10.6	24.4	-30.6	88.2
Realschule	6.9	13.8	-19.0	53.9
Integrierte Gesamtschulen	17.8	13.1	-8.5	42.7
Gymnasium	7.3	12.3	-23.9	39.9
Gesamt	8.1	15.5	-33.4	88.2

Tab. 2: Soziale Gradienten der mathematischen Kompetenz innerhalb von Schulen differenziert nach Schulformen (unstandardisierte Regressionskoeffizienten)

4.2 Sind klassenspezifische Herkunftseffekte domänenabhängig oder eher domänenübergreifend?

Die vorangegangene Analyse hat sowohl für die Lesekompetenz als auch für die mathematische Kompetenz gezeigt, dass sich Schulklassen bedeutsam darin unterscheiden, wie hoch die klassenspezifischen Disparitäten in diesen Domänen ausgeprägt sind. Dabei bleibt ungeklärt, ob die klassenspezifischen Disparitäten überwiegend domänenspezifisch oder eher domänenübergreifend sind. Um dieser Frage nachzugehen, ist in Abbildung 1 der Zusammenhang zwischen den klassenspezifischen Steigungen der Gradienten in der Lesekompetenz und in der mathematischen Kompetenz veranschaulicht. Die Korrelation zwischen den beiden Steigungsmaßen beträgt $r = .71$ und ist statistisch signifikant ($p < .05$).

Für 92 Prozent aller Schulklassen lassen sich im Lesen und in Mathematik vergleichbare soziale Disparitäten gemessen an den Gradientensteigungen feststellen. In 5 Prozent der Schulklassen fällt die Steigung der sozialen Gradienten in Mathematik höher aus als in der Lesekompetenz. In 3 Prozent der Schulklassen sind die Steigungen der sozialen Gradienten für die Lesekompetenz höher ausgeprägt als in der Mathematik. Als Kriterium für einen bedeutsamen Unterschied wurde hier eine Differenz in den unstandardisierten Gradientensteigungen von 20 Punkten gewählt, was in beiden Domänen einem Lernzuwachs von etwa einem halben Schuljahr entspricht (vgl. Schaffner,

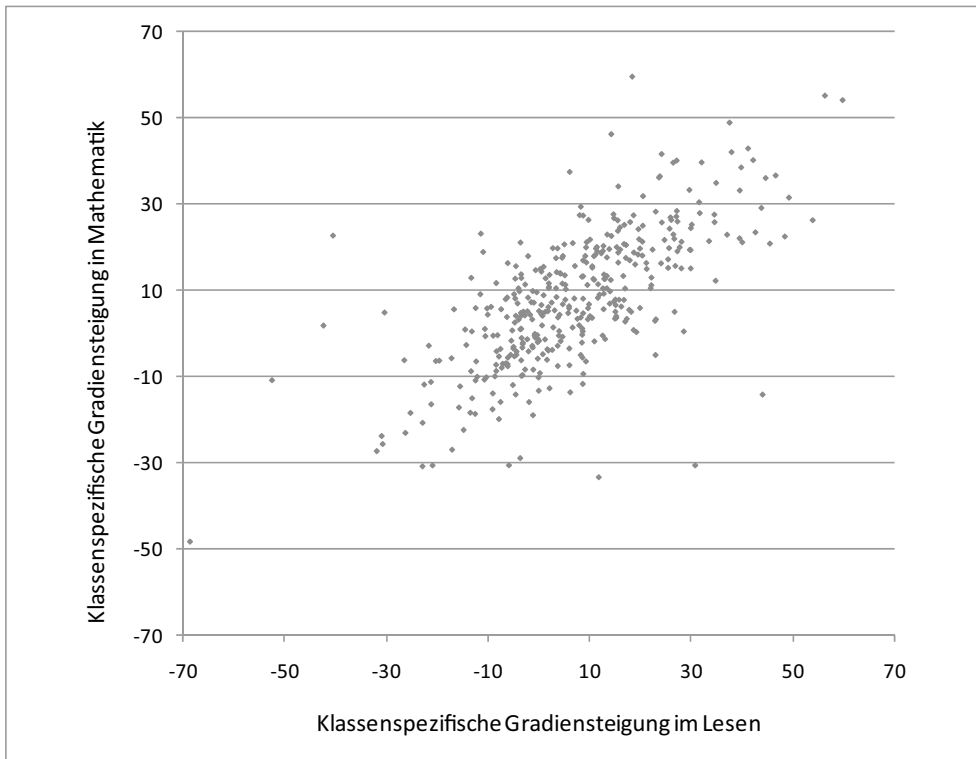


Abb. 1: Zusammenhang zwischen den klassenspezifischen Steigungen der sozialen Gradienten für die Lesekompetenz und für die mathematische Kompetenz

Schiefele, Drechsel & Artelt, 2004). Insgesamt sind die klassenspezifischen Disparitäten damit als weitgehend domänenübergreifend einzustufen.

4.3 Vorhersage von klassenspezifischen Disparitäten durch strukturelle Klassenmerkmale und unterrichtsbezogene Fördermaßnahmen

Um zu überprüfen, inwieweit strukturelle Merkmale der Klassenzusammensetzung und unterrichtsbezogene und außerschulische Förderangebote für die Ausprägung von sozialen Disparitäten innerhalb von Schulklassen prädiktiv sind, werden im Folgenden Mehrebenenanalysen durchgeführt. Dazu werden zuerst die Zusammenhänge für die klassenspezifischen Disparitäten in der Lesekompetenz analysiert (Tabelle 3) und anschließend die Ergebnisse für die mathematische Kompetenz berichtet (Tabelle 4).

Da es sich bei der vorliegenden Studie um eine Querschnittsuntersuchung handelt, lassen sich etwaige Zusammenhänge nicht direkt kausal interpretieren. Um aber zumindest eine deskriptive Abschätzung zu erlauben, wird in den folgenden Analysen das

Maß für die kognitiven Grundfähigkeiten (KFT) als Indikator für das Vorwissen kontrolliert. Zwar sagen die kognitiven Grundfähigkeiten das aktuelle Kompetenzniveau von Schülerinnen und Schülern in einer Domäne weniger gut vorher als ihr Vorwissen in derselben Domäne, doch hat sich dieses approximative Vorgehen in anderen Analysen zumindest bei der Modellierung von Kompetenzmittelwerten bewährt (vgl. Stanat, 2006; Baumert et al., 2006). Die Tabelle 3 zeigt die Ergebnisse aus vier Modellrechnungen, mit denen die Effekte von strukturellen Klassenmerkmalen und unterrichtsbezogenen und außerschulischen Förderangeboten auf das klassenspezifische Ausmaß von Disparitäten in der Lesekompetenz analysiert werden.

In dem Modell 1 sind zunächst nur die Variablen „sozioökonomischer Status (HISEI)“ und „kognitive Fähigkeiten (KFT)“ auf der Individualebene enthalten. Das Modell entspricht einer Regression ohne weitere Variablen auf der zweiten Ebene. Erwartungskonform erweisen sich auf der Individualebene sowohl der Indikator für die kognitiven Grundfähigkeiten der Schülerinnen und Schüler als auch der sozioökonomische Status der Jugendlichen als statistisch bedeutsame Prädiktoren für die Lesekompetenz. Mit einem Anstieg der kognitiven Grundfähigkeiten um eine Standardabweichung ist eine Zunahme der Lesekompetenz um 30 Punkte verbunden. Auch die Bedeutung des sozioökonomischen Status ist nachweisbar. Bei vergleichbar ausgeprägten kognitiven Grundfähigkeiten steigt die Lesekompetenz der Schülerinnen und Schüler um 7 Punkte, wenn sich der sozioökonomische Status um eine Standardabweichung erhöht.

Das Modell 2 erweitert das Modell 1 und berücksichtigt auf der Klassenebene zusätzlich Merkmale, die die strukturelle Zusammensetzung der Schülerinnen und Schüler in den Klassen beschreiben. Als Prädiktoren für die klassenspezifischen Gradientensteigungen werden das Niveau der kognitiven Fähigkeiten (klassenspezifischer Mittelwert) und das mittlere Lebensalter der Jugendlichen einbezogen. Ferner werden als Maße für heterogene Lernvoraussetzungen jeweils auf Klassenebene die Standardabweichung der kognitiven Fähigkeiten und des Lebensalters sowie der klassenbezogene Anteil von Jugendlichen mit Migrationshintergrund berücksichtigt. Das Ergebnis des Modells 2 zeigt, dass bereits bei Kontrolle der (wenigen) strukturellen Klassenmerkmale ein nennenswerter Anteil der Varianz der klassenspezifischen Gradientensteigungen aufgeklärt werden kann ($R^2_{\text{Slope}} = 0.27$). Stärkster Einzelprädiktor für hohe klassenspezifische Disparitäten in der Lesekompetenz ist die klassenspezifische Streuung in den kognitiven Fähigkeiten. Demnach sind die klassenspezifischen Disparitäten besonders hoch ausgeprägt in Schulklassen mit hohen kognitiven Leistungsunterschieden.

Das Modell 3 erweitert das Modell 2 um fünf Indikatoren zu unterrichtsbezogenen und außerschulischen Förderangeboten. Sie geben jeweils den klassenspezifischen prozentualen Anteil von Schülerinnen und Schülern an, (1) die Zusatzunterricht für leistungsstarke Jugendliche im Fach Deutsch bekommen, (2) die Zusatzunterricht für leistungsschwache Schülerinnen und Schüler im Fach Deutsch erhalten, (3) die privaten Einzelunterricht oder Einzelnachhilfe innerhalb oder außerhalb der Schule erhalten, (4) die an einer schulischen Hausaufgabenbetreuung und (5) die an einem Ganztagsangebot in der Schule teilnehmen.

Das Ergebnis des dritten Modells zeigt, dass sich die Varianzaufklärung in den klassenspezifischen Gradientensteigungen durch die Hinzunahme der prozessbezogenen Indikatoren noch einmal deutlich verbessert: 45 Prozent der Slope-Varianz können dadurch aufgeklärt werden. Spezifische und signifikante Einzeleffekte auf die klassenspezifischen Gradientensteigungen in der Lesekompetenz lassen sich aber für die berücksichtigten Merkmale nicht nachweisen. Einzige Ausnahme ist der Anteil von Jugendlichen innerhalb einer Klasse, die ein Ganztagsangebot nutzen. Je höher dieser Anteil ausfällt, desto steiler fallen die sozialen Gradienten innerhalb der Klassen aus. Dieser Befund bleibt auch dann noch stabil, wenn zusätzlich auf der Individualebene der Interaktionseffekt zwischen sozioökonomischem Status und individueller Ganztagsbeteiligung kontrolliert wird. Allerdings liegt hier vermutlich eine Konfundierung mit der Schulform „Integrierte Gesamtschule“ vor, da für diese Schulform höhere klassenspezifische Disparitäten nachgewiesen werden konnten und an Integrierten Gesamtschulen überwiegend Ganztagsangebote vorherrschen. Das Modell 4 kontrolliert daher zusätzlich noch die Schulform. Die Varianzaufklärung erhöht sich dadurch auf $R^2_{\text{Slope}} = 60$ Prozent. Wie auch schon in der deskriptiven Analyse (Tabelle 2) ergibt sich für die Integrierte Gesamtschule die stärkste klassenspezifische Gradientensteigung. Unter Kontrolle aller übrigen Merkmale liegen für diese Schulform die klassenspezifischen Steigungen der Gradienten in der Lesekompetenz bei +15 Punkten. Der Koeffizient für den Anteil an schulischen Ganztagsangeboten sinkt in diesem Modell nach Kontrolle des Interaktionseffekts auf Individualebene und der Schulform wieder unter die Signifikanzgrenze von fünf Prozent. Lediglich für die klassenspezifischen Unterschiede im Lebensalter der Schülerinnen und Schüler lässt sich ein signifikanter spezifischer Vorhersagebeitrag unter Kontrolle aller übrigen Merkmale feststellen: je altersheterogener die Klassen ausfallen, desto steiler ist der Anstieg der klassenspezifischen Gradienten der Lesekompetenz.

Zusammenfassend ergibt sich für die Vorhersage der klassenspezifischen Gradienten in der Lesekompetenz eine ambivalente Befundlage. Insgesamt können durch die berücksichtigten strukturellen und unterrichtsbezogenen Merkmale zwar bis zu 60 Prozent der Varianz in den Gradientensteigungen aufgeklärt werden. Signifikante Nachweise für die spezifische Bedeutung einzelner Merkmale sind aber über die Modelle hinweg nicht stabil. Die tendenziellen Richtungen der Vorhersagekoeffizienten sind zumindest theoretisch plausibel. So zeigen sich ausgeprägte klassenspezifische Disparitäten eher in altersheterogenen Lerngruppen und wenn hohe Anteile von leistungsstarken Schülerinnen und Schülern durch Zusatzunterricht gefördert werden. Kompensationsangebote wie Zusatzunterricht für leistungsschwache Jugendliche und schulische Hausaufgabenbetreuungen weisen hingegen tendenziell eher in Richtung geringerer klasseninterner Disparitäten. Die spezifischen Koeffizienten bleiben allerdings unter dem Signifikanzniveau.

	Modell 1		Modell 2		Modell 3		Modell 4	
	B	SE	B	SE	B	SE	B	SE
Schätzung der Klassenmittelwerte								
Intercept	491.9	(2.5)	511.9	(8.6)	510.6	(8.1)	478.8	(9.1)
Kognitive Fähigkeiten (MW)			51.2	(3.5)	48.4	(3.4)	30.9	(4.3)
Kognitive Fähigkeiten (SD)			-2.3	(8.1)	-1.6	(7.5)	4.8	(7.5)
Alter (MW)			-18.5	(4.0)	-16.3	(4.0)	-14.1	(3.9)
Alter (SD)			-11.7	(7.6)	-9.8	(7.2)	-5.9	(6.8)
Migrationshintergrund (in %)			-24.4	(10.5)	-21.5	(9.6)	-25.0	(9.9)
Zusatzunterricht für leistungsstarke Schüler (Deutsch), (in %)					-8.7	(30.5)	8.9	(27.6)
Zusatzunterricht für leistungsschwache Schüler (Deutsch), (in %)					-59.2	(25.8)	-47.9	(24.4)
Teilnahme an privatem Einzelunterricht (in %)					35.1	(14.3)	10.6	(14.3)
Teilnahme an schulischer Hausaufgabenbetreuung (in %)					-26.5	(24.5)	-13.8	(24.3)
Teilnahme an schulischem Ganztagsangebot (in %)					-5.1	(8.4)	-7.6	(8.4)
Schulformen (Referenzgruppe Hauptschule)								
Schulen mit mehreren Bildungsgängen							14.6	(5.2)
Realschule							24.4	(5.9)
Integrierte Gesamtschulen							15.8	(5.1)
Gymnasien							45.8	(7.1)
Schätzung der HISEI-Steigung								
Intercept	6.7	(0.9)	-7.2	(5.3)	-9.6	(5.5)	-11.2	(6.2)
Kognitive Fähigkeiten (MW)			2.0	(2.9)	2.2	(2.9)	0.0	(3.3)
Kognitive Fähigkeiten (SD)			10.0	(4.8)	8.2	(4.8)	1.9	(4.7)
Alter (MW)			-2.0	(2.8)	-3.1	(2.7)	-1.8	(2.8)
Alter (SD)			6.4	(4.8)	8.3	(4.8)	9.7	(4.6)
Migrationshintergrund (in %)			-1.8	(6.4)	-0.7	(5.9)	-1.3	(5.7)
Zusatzunterricht für leistungsstarke Schüler (Deutsch), (in %)					15.1	(13.7)	22.3	(13.1)
Zusatzunterricht für leistungsschwache Schüler (Deutsch), (in %)					-6.5	(13.7)	-8.9	(12.8)
Teilnahme an privatem Einzelunterricht (in %)					5.8	(9.9)	2.6	(9.4)
Teilnahme an schulischer Hausaufgabenbetreuung (in %)					-27.7	(17.6)	-25.9	(16.9)
Teilnahme an schulischem Ganztagsangebot (in %)					16.5	(7.1)	12.6	(6.7)
Schulformen (Referenzgruppe Hauptschule)								
Schulen mit mehreren Bildungsgängen							5.8	(4.8)
Realschule							4.9	(3.7)
Integrierte Gesamtschulen							15.0	(4.3)
Gymnasien							5.8	(4.8)
Kontrolle von KFT								
Kognitive Fähigkeiten	29.6	(0.9)	27.1	(0.9)	27.1	(0.9)	27.0	(0.9)
Kontrolle von Interaktionen								
Interaktion HISEI x Schulisches Ganztagsangebot					2.4	(3.3)	2.3	(3.3)
Varianzaufklärung								
Intercept			0.82		0.84		0.87	
Slope			0.27		0.45		0.60	

Anmerkung: Fettgedruckte Koeffizienten sind signifikant (p < .05).

Tab. 3: Effekte von strukturellen Merkmalen der Klassenzusammensetzung und von Indikatoren für kompensatorische Schul- und Unterrichtsmerkmale auf die klassenspezifischen Disparitäten in der Lesekompetenz (unstandardisierte Regressionskoeffizienten aus Mehrebenenanalysen)

Die Tabelle 4 zeigt die Ergebnisse für die klassenspezifischen Disparitäten in der mathematischen Kompetenz. Die vier Modellrechnungen wurden analog zu den Analysen in der Lesekompetenz durchgeführt. Die Prädiktorvariablen sind weitgehend identisch. Lediglich die Angaben hinsichtlich des Zusatzunterrichts für leistungsschwache bzw. leistungsstarke Schülerinnen und Schüler beziehen sich jetzt auf den Mathematikunterricht (und nicht mehr auf den Deutschunterricht).

Fixed Effects	Modell 1		Modell 2		Modell 3		Modell 4	
	B	SE	B	SE	B	SE	B	SE
Schätzung der Klassenmittelwerte								
Intercept	504.8	(2.3)	527.6	(9.0)	528.0	(8.8)	493.4	(9.2)
Kognitive Fähigkeiten (MW)			42.6	(3.8)	40.4	(3.7)	22.5	(4.2)
Kognitive Fähigkeiten (SD)			-11.8	(9.1)	-9.4	(8.9)	2.1	(8.4)
Alter (MW)			-16.0	(4.0)	-15.2	(3.8)	-11.5	(3.5)
Alter (SD)			-6.4	(6.8)	-6.6	(6.3)	-0.7	(5.9)
Migrationshintergrund (in %)			-28.9	(8.3)	-24.5	(8.0)	-29.6	(8.2)
Zusatzunterricht für leistungsstarke Schüler (Mathematik) (in %)					-33.9	(16.1)	-13.2	(14.4)
Zusatzunterricht für leistungsschwache Schüler (Mathematik) (in %)					0.5	(11.5)	-6.1	(10.3)
Teilnahme an privatem Einzelunterricht (in %)					17.3	(15.3)	-3.2	(15.1)
Teilnahme an schulischer Hausaufgabenbetreuung (in %)					-6.9	(21.4)	12.7	(21.1)
Teilnahme an schulischem Ganztagsangebot (in %)					-5.9	(8.6)	-7.7	(7.7)
Schulformen (Referenzgruppe Hauptschule)								
Schulen mit mehreren Bildungsgängen							4.8	(5.5)
Realschule							17.8	(5.1)
Integrierte Gesamtschulen							7.2	(5.0)
Gymnasien							47.2	(6.3)
Schätzung der HISEI-Steigung								
Intercept	8.2	(1.0)	-1.2	(5.5)	-2.3	(6.1)	-4.5	(7.1)
Kognitive Fähigkeiten (MW)			4.4	(2.5)	5.1	(2.8)	2.2	(3.3)
Kognitive Fähigkeiten (SD)			9.5	(5.2)	7.9	(5.0)	4.0	(5.0)
Alter (MW)			-0.4	(2.6)	-1.3	(2.6)	0.5	(2.5)
Alter (SD)			0.5	(4.4)	1.9	(4.4)	2.6	(4.5)
Migrationshintergrund (in %)			3.3	(5.3)	2.8	(5.4)	2.0	(5.4)
Zusatzunterricht für leistungsstarke Schüler (Mathematik) (in %)					13.0	(10.6)	17.9	(10.9)
Zusatzunterricht für leistungsschwache Schüler (Mathematik) (in %)					-5.4	(7.3)	-6.7	(7.1)
Teilnahme an privatem Einzelunterricht (in %)					-0.3	(9.7)	-3.3	(9.9)
Teilnahme an schulischer Hausaufgabenbetreuung (in %)					-12.3	(14.0)	-9.0	(13.0)
Teilnahme an schulischem Ganztagsangebot (in %)					11.8	(6.6)	10.2	(5.7)
Schulformen (Referenzgruppe Hauptschule)								
Schulen mit mehreren Bildungsgängen							3.3	(4.4)
Realschule							4.0	(3.6)
Integrierte Gesamtschulen							8.1	(3.6)
Gymnasien							6.3	(4.8)
Kontrolle von KFT								
Kognitive Fähigkeiten	37.5	(1.0)	34.9	(1.0)	34.9	(1.0)	34.8	(1.0)
Kontrolle von Interaktionen								
Interaktion HISEI x Schulisches Ganztagsangebot					0.2	(2.8)		
Varianzaufklärung Intercept			0.79		0.80		0.86	
Varianzaufklärung Slope			0.08		0.25		0.33	

Anmerkung: Fettgedruckte Koeffizienten sind signifikant ($p < .05$).

Tab. 4: Effekte von strukturellen Merkmalen der Klassenzusammensetzung und von Indikatoren für kompensatorische Schul- und Unterrichtsmerkmale auf die klassenspezifischen Disparitäten in der mathematischen Kompetenz (unstandardisierte Regressionskoeffizienten aus Mehrebenenanalysen)

Die Ergebnisse für die klassenspezifischen Disparitäten in der mathematischen Kompetenz bestätigen im Wesentlichen die Ergebnisse für die Lesekompetenz. Auch für die Domäne Mathematik gibt es nachweisbare Disparitäten innerhalb der Schulklassen. Die klassenspezifischen Gradientensteigungen betragen hier durchschnittlich 8 Punkte (Tabelle 4, Modell 1). Wie auch für die Lesekompetenz lassen sich für die einzelnen strukturellen und unterrichtsbezogenen Prädiktoren statistisch im Wesentlichen keine

spezifischen Vorhersagebeiträge absichern. Die tendenziellen Richtungen der Prädiktoren sind aber auch hier theoretisch plausibel. So steigen die klassenspezifischen Gradienten mit zunehmender Förderung von leistungsstarken Schülerinnen und Schülern und in eher altersheterogenen Lerngruppen. Hingegen fallen die klasseninternen Disparitäten in der Mathematik tendenziell geringer aus, wenn leistungsschwache Schülerinnen und Schüler Förderunterricht erhalten oder wenn viele Jugendliche einer Schulklasse an einer schulischen Hausaufgabenbetreuung teilnehmen.

Nicht erwartungskonform ist wiederum die tendenziell positive Richtung des Koeffizienten für die Teilnahme an schulischen Ganztagsangeboten. Die klasseninternen Disparitäten in der mathematischen Kompetenz fallen demnach stärker aus, wenn ein hoher Anteil der Schülerinnen und Schüler in einer Klasse an einem Ganztagsschulprogramm teilnimmt.

Einen wichtigen Unterschied zu den Modellrechnungen in der Lesekompetenz liefert der Vergleich der Varianzaufklärungen. Die Varianz in der klassenspezifischen Gradientensteigung in der Mathematik lässt sich bis zu 33 Prozent durch die einbezogenen Prädiktoren aufklären. Der Wert fällt damit nur etwa halb so hoch aus wie für die Lesekompetenz, bei der die Varianzaufklärung bei 60 Prozent lag. Demnach sind die berücksichtigten strukturellen und unterrichtsbezogenen Merkmale für die klassenspezifischen Disparitäten in der Lesekompetenz vorhersagekräftiger als für die mathematische Kompetenz.

5. Zusammenfassung und Diskussion

Ziel dieses Beitrags war es, die sozialen Disparitäten im Lesen und in der Mathematik innerhalb von Klassen zu beschreiben und mögliche Erklärungsfaktoren zu identifizieren. Als Stichprobe dienten 393 Schulklassen mit mehr als 9000 Schülerinnen und Schülern der neunten Jahrgangsstufe, die im Rahmen von PISA 2009 getestet wurden.

Die deskriptiven Befunde zeigten, dass die klassenspezifischen Disparitäten in der neunten Jahrgangsstufe in Deutschland eher gering ausgeprägt sind. Die Steigungen der sozialen Gradienten innerhalb der Klassen betragen durchschnittlich 7 Punkte im Lesen und 8 Punkte in Mathematik. Ein wichtiges Ergebnis waren aber die erheblichen Unterschiede in den klassenspezifischen Disparitäten zwischen den Klassen. Die Steigungen der sozialen Gradienten variierten in einem Wertebereich zwischen –53 und 74 Punkten im Lesen und zwischen –33 und 88 Punkten in der Mathematik. Die klassenspezifischen sozialen Disparitäten sind also nur in einem Teil der Schulklassen besonders stark ausgeprägt, in anderen Schulklassen finden sich hingegen sogar „umgekehrte“ Herkunftseffekte.

Eine nach Schulformen getrennte Auswertung ergab, dass die höchsten klassenspezifischen sozialen Disparitäten an den Integrierten Gesamtschulen und den Schulen mit mehreren Bildungsgängen bestehen. Die klassenspezifischen Disparitäten an den Hauptschulen, Realschulen und Gymnasien waren dagegen geringer ausgeprägt. Damit scheinen die sozialen Disparitäten innerhalb der Klassen höher auszufallen, wenn es

sich um Schulformen handelt, deren Schülerschaft generell heterogener hinsichtlich der schulischen Leistungen und der sozialen Zusammensetzung ist.

Im Hinblick auf die Frage, ob die klassenspezifischen Disparitäten eher domänenübergreifend oder eher domänenabhängig auftreten, zeigte sich ein enger Zusammenhang zwischen beiden Domänen ($r = .71$). Nur in weniger als 10 Prozent der Klassen unterschieden sich die Gradientensteigungen im Lesen und in der Mathematik bedeutsam. Eine eindeutige Tendenz für eines der beiden Fächer war nicht zu erkennen. Auch in der mittleren Höhe waren die Steigungen der klassenspezifischen Gradienten im Lesen und in der Mathematik vergleichbar. Klassenspezifische Herkunftseffekte lassen sich also weitgehend in beiden Domänen gleichermaßen feststellen. Dies bestätigen auch die Befundmuster von Ma (2000, 2001).

Der Frage, welche Bedingungsfaktoren für klassenspezifische Disparitäten prädiktiv sind, wurde in einem dritten Auswertungsschritt anhand von Mehrebenenanalysen nachgegangen. In beiden Domänen konnte ein bedeutsamer Anteil in der Varianz der Gradientensteigungen zwischen den Schulklassen aufgeklärt werden. Die Varianzaufklärung fiel im Lesen ($R^2_{\text{Slope}} = 0.60$) etwa doppelt so hoch aus wie in der Mathematik ($R^2_{\text{Slope}} = 0.31$). Obwohl das Niveau der klassenspezifischen Gradientensteigungen in beiden Domänen vergleichbar war, scheinen die strukturellen und unterrichtsbezogenen Merkmale für die klassenspezifischen Disparitäten in der Lesekompetenz vorhersagekräftiger als für die mathematische Kompetenz. In der Mathematik scheinen offenbar noch andere Prozesse und Bedingungsfaktoren relevant zu sein. Inwieweit naturwissenschaftsbezogene schulische Nachmittagsangebote und Unterschiede in der elterlichen Wertschätzung von Mathematik und Naturwissenschaften hier eine Rolle spielen, sollte in weiteren Studien noch untersucht werden.

Ferner wurde der Frage nachgegangen, inwieweit einzelne strukturelle Merkmale oder unterrichtsbezogene Faktoren prädiktiv für klassenspezifische Disparitäten in den beiden Domänen waren. Hier ließen sich für die meisten Einzelprädiktoren statistisch keine spezifischen Vorhersagebeiträge absichern. Die tendenziellen Richtungen der Prädiktoren waren aber weitgehend theoretisch plausibel. Die klassenspezifischen Gradienten steigen der Tendenz nach bei zunehmender Förderung von leistungsstarken Schülerinnen und Schülern und in eher altersheterogenen Lerngruppen an. Die Disparitäten innerhalb von Klassen fallen hingegen geringer aus, wenn leistungsschwache Schülerinnen und Schüler mehr Förderunterricht erhalten oder wenn viele Jugendliche an einer schulischen Hausaufgabenbetreuung teilnehmen.

Ein unerwarteter Befund ergab sich für das Merkmal „Teilnahme an Ganztagsangeboten“. In Klassen, in denen ein hoher Anteil von Schülerinnen und Schülern an einem schulischen Ganztagsangebot teilnimmt, zeigten sich bei Kontrolle der übrigen Merkmale höhere klassenspezifische Disparitäten im Lesen als in Schulklassen, in denen nur wenige Jugendliche an einem Ganztagsangebot teilnehmen. Dieses Ergebnis könnte inhaltlich dahingehend interpretiert werden, dass möglicherweise die inhaltlichen Angebote an Ganztagschulen abhängig von der sozialen Herkunft unterschiedlich genutzt werden. Unter Umständen wählen Kinder aus bildungsnahen Schichten hier häufiger Angebote aus, die zusätzliche Lerngelegenheiten bieten und sich insgesamt positiv

auf die schulischen Leistungen auswirken, während Jugendliche aus eher bildungsfernen Schichten häufiger andere Angebote annehmen. Diese Vermutung muss allerdings noch durch weitere Studien geprüft werden, denn in den komplexesten Modellrechnungen (unter Kontrolle der Schulform) ließ sich statistische Signifikanz nicht mehr absichern. Hier liegt vermutlich eine Konfundierung mit der Schulform „Integrierte Gesamtschule“ vor, an der überwiegend Ganztagsangebote vorherrschen.

Eine methodische Erklärung dafür, dass sich die Vorhersageeffekte der strukturellen und unterrichtsbezogenen Klassenmerkmale überwiegend nicht statistisch absichern ließen, liegt möglicherweise auch in grundsätzlichen Problemen beim Schätzen von Slope-As-Outcome-Modellen (Raudenbush & Bryk, 2002, S. 117). So haben Regressionskoeffizienten eine höhere „sampling variability“ als Mittelwerte einer Stichprobe. Die Schätzung von Regressionskoeffizienten in einer kleinen Stichprobe ist mit hohen Standardschätzfehlern behaftet. Die geringe Reliabilität der Schätzungen der Gradientensteigungen führt dazu, dass sich Vorhersageeffekte auch bei einem Datensatz mit immerhin 393 Schulklassen nur schwer nachweisen lassen.

Zusammenfassend kann man im Hinblick auf die eingangs formulierten theoretischen Vermutungen festhalten: sowohl strukturelle Klassenmerkmale als auch unterrichtsbezogene und außerschulische Fördermaßnahmen tragen bedeutsam zur Vorhersage von klassenspezifischen Disparitäten im Lesen und in der Mathematik bei. Die vermuteten kompensatorischen Effekte auf die klasseninternen Disparitäten für einzelne Interventionsmaßnahmen wie etwa die Teilnahme am schulischen Förderunterricht oder an einer Hausaufgabenbetreuung zeigten sich aber nur der Tendenz nach. Vor dem Hintergrund aktueller schulstruktureller Entwicklungen wie dem Zusammenlegen von mehreren Bildungsgängen in mehreren Bundesländern zu neuen Schulformen ist mit einem Anstieg der Heterogenität innerhalb der Schulen zu rechnen. Die leistungs-, alters- und herkunftsbezogene Streuung innerhalb der Schulen und Klassen wird voraussichtlich ansteigen, entsprechend der gezeigten Befunde werden sich damit möglicherweise auch die sozialen Disparitäten innerhalb der Schulen erhöhen. Für die pädagogische Arbeit an Schulen wird dies eine zunehmende Herausforderung bedeuten, aber zugleich auch die Chance, durch Interventionsmaßnahmen und schulische Angebote solche sozialen Disparitäten abzuschwächen.

Mit diesen Analysen konnte somit der Beitrag geleistet werden, auf die Bedeutung von sozialen Disparitäten innerhalb von Schulen hinzuweisen und mögliche Anhaltspunkte für deren Entstehung zu identifizieren. Allerdings müssen diese Ergebnisse noch in weiteren Studien validiert und ausgebaut werden. Das hier verwendete Datenmaterial ist aufgrund der querschnittlichen Anlage nur begrenzt geeignet, die kausalen Effekte bei der Entstehung von klassenspezifischen Disparitäten abzubilden. Schließlich sind die berichteten Effekte auch immer Ergebnisse kumulativer Prozesse über die Schulzeit hinweg. Die Schülerinnen und Schüler wurden in ihrer neunjährigen Schullaufbahn von einer größeren Anzahl von Lehrkräften unterrichtet und haben über die Zeit wechselnde Klassenverbände innerhalb einer Schule besucht. Inwieweit es sich bei den Befunden daher tatsächlich um klassenspezifische Effekte handelt oder nicht doch eher um schulspezifische Effekte, kann mit dem vorliegenden Datensatz nicht geklärt werden.

Abschließend sollen noch zwei Empfehlungen für weitere Forschungen in diesem Bereich gegeben werden: (1.) Um die Entstehung von sozialen Disparitäten innerhalb von Klassen weiter zu erforschen, wäre es wichtig, auch jüngere Alterskohorten etwa in der Grundschule in den Blick zu nehmen, da hier vermutlich aufgrund der heterogenen Schülerschaft die sozialen Disparitäten innerhalb der Klassen noch eine größere Rolle spielen als in der Sekundarstufe I. (2.) Vom methodischen Vorgehen könnte es außerdem weiterführend sein, nicht nur die Bedeutung spezifischer Einzelmerkmale auf die Entstehung von klassenspezifischen Disparitäten zu untersuchen, sondern ganze Merkmalsprofile von Schulklassen in den Blick zu nehmen. Als statistische Methode würden sich dafür etwa latente Klassenanalysen anbieten. Dabei könnten auf der Ebene von Schulklassen Profile von Schul- und Unterrichtsmerkmalen identifiziert werden, die mit hohen oder bestenfalls geringen klassenspezifischen Disparitäten einhergehen.

Literatur

- Baumert, J., Stanat, P., & Watermann, R. (2006). Schulstruktur und die Entstehung differenzieller Lern- und Entwicklungsmilieus. In J. Baumert, P. Stanat & R. Watermann (Hrsg.), *Herkunftsbedingte Disparitäten im Bildungswesen. Vertiefende Analysen im Rahmen von PISA 2000* (S. 95-188). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Becker, M., Stanat, P., Baumert, J., & Lehmann, R. H. (2008). Lernen ohne Schule: Differentielle Entwicklung der Leseleistungen von Kindern mit und ohne Migrationshintergrund während der Sommerferien. In F. Kalter (Hrsg.), *Migration und Integration* (Kölner Zeitschrift für Soziologie und Sozialpsychologie, Sonderheft Nr. 48, S. 252-276). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Bos, W., Dohe, C., Kowoll, M. E., Platz, U., & Schuster, S. (2010). Binnendifferenzierung als integrationspädagogische Maßnahme. In M. Krüger-Potratz, U. Neumann & H. H. Reich (Hrsg.), *Bei Vielfalt Chancengleichheit. Interkulturelle Pädagogik und Durchgängige Sprachbildung* (S. 296-305). Münster: Waxmann.
- Bos, W., & Scharenberg, K. (2010). Lernentwicklung in leistungshomogenen und -heterogenen Schulklassen. In W. Bos, E. Klieme & O. Köller (Hrsg.), *Schulische Lerngelegenheiten und Kompetenzentwicklung* (S. 173-194). Münster: Waxmann.
- Boudon, R. (1974). *Education, opportunity, and social inequality: Changing prospects in Western society*. New York: Wiley.
- Bourdieu, P. (1982). *Die feinen Unterschiede. Kritik der gesellschaftlichen Urteilskraft*. Frankfurt a.M.: Suhrkamp.
- Bourdieu, P., & Passeron, J.-C. (1971). *Die Illusion der Chancengleichheit. Untersuchungen zur Soziologie des Bildungswesens am Beispiel Frankreichs*. Stuttgart: Klett.
- Cooper, H., Nye, B., Charlton, K., Lindsay, J., & Greathouse, S. (1996). The effects of summer vacation on achievement test scores: A narrative and meta-analytic review. *Review of Educational Research*, 66(3), 227-268.
- Ditton, H. (2010). Wie viel Ungleichheit durch Bildung verträgt eine Demokratie? *Zeitschrift für Pädagogik*, 56(1), 53-68.
- Ditton, H., Krüskens, J., & Schauenberg, M. (2005). Bildungsungleichheit – der Beitrag von Familie und Schule. *Zeitschrift für Erziehungswissenschaft*, 8(2), 285-303.
- Ehmke, T. (2008). Welche Bedeutung haben lernförderliche und naturwissenschaftsbezogene Einstellungen und Prozesse im Elternhaus für die Erklärung sozialer Disparitäten in der naturwissenschaftlichen Kompetenz? *Zeitschrift für Erziehungswissenschaft, Sonderheft 10*, 129-148.

- Ehmke, T., & Jude, N. (2010). Soziale Herkunft und Kompetenzerwerb. In E. Klieme, C. Artelt, J. Hartig, N. Jude, O. Köller, M. Prenzel, W. Schneider & P. Stanat (Hrsg.), *PISA 2009. Bilanz nach einem Jahrzehnt* (S. 231-254). Münster: Waxmann.
- Ehmke, T., & Siegle, T. (2005). ISEL, ISCED, HOMEPOS, ESCS. Indikatoren der sozialen Herkunft bei der Quantifizierung von sozialen Disparitäten. *Zeitschrift für Erziehungswissenschaft*, 8(4), 521-540.
- Ehmke, T., & Siegle, T. (2008). Einfluss elterlicher Mathematikkompetenz und familialer Prozesse auf den Kompetenzerwerb von Kindern in Mathematik. *Psychologie in Erziehung und Unterricht*, 55(4), 253-264.
- Ganzeboom, H. B. G., de Graaf, P. M., Treiman, D. J., & de Leeuw, J. (1992). A standard international socio-economic index of occupational status. *Social Science Research*, 21(1), 1-56.
- Hattie, J. A. C. (2002). Classroom composition and peer effects. *International Journal of Educational Research*, 37(5), 449-481.
- Heller, K. A., & Perleth, C. (2000). *KFT 4-12+R – Kognitiver Fähigkeitstest für 4. bis 12. Klassen, Revision*. Göttingen: Beltz.
- International Labor Office (1990). *International standard classification of occupations. ISCO-88*. Genf: International Labor Office.
- Jude, N., & Klieme, E. (2010). Das Programme for International Student Assessment (PISA). In E. Klieme, C. Artelt, J. Hartig, N. Jude, O. Köller, M. Prenzel, W. Schneider & P. Stanat (Hrsg.), *PISA 2009. Bilanz nach einem Jahrzehnt* (S. 11-22). Münster: Waxmann.
- Klieme, E., Artelt, C., Hartig, J., Jude, N., Köller, O., Prenzel, M., Schneider, W., & Stanat, P. (2010). *PISA 2009. Bilanz nach einem Jahrzehnt*. Münster: Waxmann.
- Lehmann, R. H., Peek, R., & Gänsfuß, R. (1997). *Aspekte der Lernausgangslage von Schülerinnen und Schülern der fünften Klassen an Hamburger Schulen*. Hamburg: Behörde für Schule, Jugend und Berufsbildung, Amt für Schule.
- Lüdtke, O., Robitzsch, A., & Köller, O. (2002). Statistische Artefakte bei Kontexteffekten in der pädagogisch-psychologischen Forschung. *Zeitschrift für Pädagogische Psychologie*, 16(3/4), 217-231.
- Ma, X. (2000). Socioeconomic gaps in academic achievement within schools: Are they consistent across subject areas? *Educational Research and Evaluation*, 6, 337-356.
- Ma, X. (2001). Stability of socioeconomic gaps in mathematics and science achievement among Canadian schools. *Canadian Journal of Education*, 26(1), 97-118.
- Ma, X., & Klinger, D. A. (2000). Hierarchical linear modelling of student and school effects on academic achievement. *Canadian Journal of Education*, 25(1), 41-55.
- Maaz, K., Baumert, J., & Trautwein, U. (2009). Genese sozialer Ungleichheit im institutionellen Kontext der Schule: Wo und wie vergrößert sich soziale Ungleichheit? *Zeitschrift für Erziehungswissenschaften, Sonderheft 12*, 11-44.
- Maaz, K., & Nagy, G. (2009). Der Übergang von der Grundschule in die weiterführenden Schulen des Sekundarschulsystems: Definition, Spezifikation und Quantifizierung primärer und sekundärer Herkunftseffekte. *Zeitschrift für Erziehungswissenschaft, Sonderheft 12*, 153-182.
- Maaz, K., Trautwein, U., Lüdtke, O., & Baumert, J. (2008). Educational Transitions and Differential Learning Environments: How Explicit Between-School Tracking Contributes to Social Inequality in Educational Outcomes. *Child Development Perspectives*, 2(2), 99-106.
- OECD (2009). *PISA 2009. Assessment Framework. Key competencies in reading, mathematics and science*. Paris: OECD.
- OECD (2010). *Overcoming social background. Equity in learning opportunities and outcomes*. Paris: OECD.
- OECD (2012). *PISA 2009 Technical Report (Preliminary version)*. Paris: OECD.
- Opdenakker, M.-C., & van Damme, J. (2001). Relationship between school composition and characteristics of school process and their effect on mathematics achievement. *British Educational Research Journal*, 27(4), 407-432.

- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models. Applications and data analysis methods*. Thousand Oaks: Sage.
- Raudenbush, S. W., Bryk, A. S., & Congdon, R. (2004). *HLM 6: Hierarchical Linear & Non-linear Modeling*. Lincolnwood: SSI Scientific Software International.
- Schafer, J. L. (2000). *Norm for Windows 95/98/NT* (Version 2.03).
- Schafer, J. L., & Graham, J. W. (2002). Missing Data: Our View of the State of the Art. *Psychological Methods*, 7(2), 147-177.
- Schaffner, E., Schiefele, U., Drechsel, B., & Artelt, C. (2004). Lesekompetenz. In M. Prenzel, J. Baumert, W. Blum, R. H. Lehmann, D. Leutner, M. Neubrand, R. Pekrun, H.-G. Rolff, J. Rost & U. Schiefele (Hrsg.), *PISA 2003. Der Bildungsstand der Jugendlichen in Deutschland – Ergebnisse des zweiten internationalen Vergleichs* (S. 93-110). Münster: Waxmann.
- Stanat, P. (2006). Disparitäten im schulischen Erfolg: Analysen zur Rolle des Migrationshintergrunds. *Unterrichtswissenschaft*, 43, 98-124.
- Stanat, P. (2006). Schulleistungen von Jugendlichen mit Migrationshintergrund: Die Rolle der Zusammensetzung der Schülerschaft. In J. Baumert, P. Stanat & R. Watermann (Hrsg.), *Herkunftsbedingte Disparitäten im Bildungswesen: Differenzielle Bildungsprozesse und Probleme der Verteilungsgerechtigkeit* (S. 189-219). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Steuer, G., Berner, V.-D., & Dresel, M. (2010). *Moderiert der Unterricht die Stärke des Zusammenhangs zwischen Herkunft und Schulleistung? Ergebnisse von Random-Slope und Slope-As-Outcome-Analysen* (Vortrag auf der 74. Tagung der Arbeitsgruppe für Empirische Pädagogische Forschung (AEPF)).
- Thrupp, M., Lauder, H., & Robinson, T. (2002). School composition and peer effects. *International Journal of Educational Research*, 37(5), 483-504.

Anschrift des Autors

Prof. Dr. Timo Ehmke, Leuphana Universität Lüneburg, Fakultät I Institut für Bildungswissenschaft,
Scharnhorststraße 1, 21335 Lüneburg, Deutschland
E-Mail: tehmke@leuphana.de

Schülerinnen und Schüler mit Migrationshintergrund

Aileen Edele/Petra Stanat/Susanne Radmann/Michael Segeritz

Kulturelle Identität und Lesekompetenz von Jugendlichen aus zugewanderten Familien

1. Einleitung

Identitätsentwicklung wird allgemein als zentrale Entwicklungsaufgabe des Jugendalters angesehen (z.B. Erikson, 1968; Fend, 2005; Marcia, Waterman, Matteson, Archer & Orłowsky, 1993). Jugendliche aus zugewanderten Familien stehen dabei vor der besonderen Herausforderung, eine kulturelle Identität zu entwickeln, die sich einerseits auf den Herkunftskontext ihrer Familie und andererseits auf das Einwanderungsland beziehen kann (Berry, Phinney, Kwak & Sam, 2006). In der Literatur finden sich zudem Hinweise darauf, dass die kulturelle Identität von Heranwachsenden mit Zuwanderungshintergrund mit ihrer schulischen Adaption (z.B. schulbezogene Einstellungen, Schulerfolg) zusammenhängt (z.B. Berry, Phinney, Sam & Vedder, 2006a; Horenczyk, 2010). In Deutschland wurden Muster der kulturellen Identität von Jugendlichen aus zugewanderten Familien und deren Zusammenhang zu schulischer Anpassung bislang jedoch kaum untersucht.

1.1 Akkulturation und kulturelle Identität

Wenn Menschen aus unterschiedlichen kulturellen Kontexten längerfristig miteinander in Kontakt treten, setzen Anpassungsprozesse ein, die als Akkulturation bezeichnet werden (z.B. Berry, 1997). Dabei sind kollektive Veränderungen in den alltäglichen Gewohnheiten einer Gruppe zu unterscheiden von individuellen, innerpsychischen Veränderungen in den Einstellungen und der kulturellen Identität von Personen, die als psychologische Akkulturation bezeichnet werden (Berry, Phinney, Sam & Vedder, 2006b; Graves, 1967).

Nach Berry (1980; 1997) lassen sich anhand von zwei theoretisch unabhängigen Dimensionen – der Orientierung an der Herkunftskultur und der Orientierung an der Aufnahmekultur – vier verschiedene Akkulturationsorientierungen voneinander abgrenzen (vgl. Abbildung 1). Eine starke Orientierung an der Herkunftskultur bei gleichzeitig

schwacher Orientierung an der Aufnahmekultur wird als *Separation* bezeichnet, eine starke Orientierung an der Aufnahmekultur bei gleichzeitig schwacher Orientierung an der Herkunftskultur als *Assimilation*. *Integration* ist durch eine starke Orientierung an beiden Kulturen gekennzeichnet, wohingegen eine schwache Orientierung an beiden Kulturen als *Marginalisierung* bezeichnet wird. Der Begriff Marginalisierung wurde allerdings dahingehend kritisiert, dass eine schwache Orientierung an der Herkunft- und Aufnahmekultur nicht zwingend mit gesellschaftlicher Ausgrenzung einhergehe, wie von der Bezeichnung impliziert (Maehler, 2012; Rudmin & Ahmadzadeh, 2001). Infolgedessen wurde vorgeschlagen, eine schwache Orientierung an beiden Kulturen entweder weiter zu differenzieren (Bourhis, Moise, Perreault & Senécal, 1997) oder neutraler zu bezeichnen, beispielsweise als *indifferent* (Maehler, 2012). Um die Anschlussfähigkeit an den internationalen Forschungsstand zu gewährleisten, wird hier jedoch die geläufigste Bezeichnung, also Marginalisierung, verwendet.

Die kulturelle Identität von Einwanderern und ihren Nachkommen kann als ein Teilaspekt der Akkulturation, genauer gesagt der psychologischen Akkulturation, aufgefasst werden (z.B. Phinney, 1990; Phinney, Berry, Vedder & Liebkind, 2006).¹ Identität – auch kulturelle Identität – ist ein komplexes Konstrukt, das in verschiedenen Forschungstraditionen unterschiedlich konzeptualisiert und definiert wird. Häufig wird kulturelle Identität im Sinne der sozialen Identitätstheorie (Lewin, 1948; Tajfel & Turner, 1979) als das Zugehörigkeitsgefühl zu einer sozialen Gruppe aufgefasst (z.B. Horenczyk, 2008; Phinney, 1990, 1992). Danach beruht das Selbstbild nicht nur auf der persönlichen Identität, sondern auch auf dem Wissen, einer sozialen Gruppe (oder mehreren sozialen Gruppen) anzugehören, sowie auf dem Wert und der Bedeutung, die das Individuum dieser Gruppenzugehörigkeit beimisst. Die kulturelle Identität gilt als zentral für die Ausbildung des Selbstbildes, aber auch für die psychische und psychosoziale Anpassung von Jugendlichen mit Zuwanderungshintergrund (vgl. z.B. Berry et al., 2006a; Phinney, 1990).

Analog zur Akkulturation kann auch kulturelle Identität zweidimensional konzeptualisiert werden, nämlich anhand der Identifikation mit der Herkunftsgruppe einerseits und anhand der Identifikation mit der Gruppe des Aufnahmekontextes andererseits (Phinney, Horenczyk, Liebkind & Vedder, 2001; Phinney & Ong, 2007). Entsprechend können, wie bei den oben beschriebenen Akkulturationsorientierungen, die vier (kulturellen) Identitätsorientierungen Integration, Assimilation, Separation und Marginalisierung unterschieden werden (vgl. Phinney, 1990). Diese sind in Abbildung 1 dargestellt.

1 Die Forschungslinien zu Akkulturationsorientierungen und kultureller Identität haben sich zwar weitgehend unabhängig voneinander entwickelt, wurden in der Literatur aber bereits mehrfach aufeinander bezogen (z.B. Berry et al., 2006a).

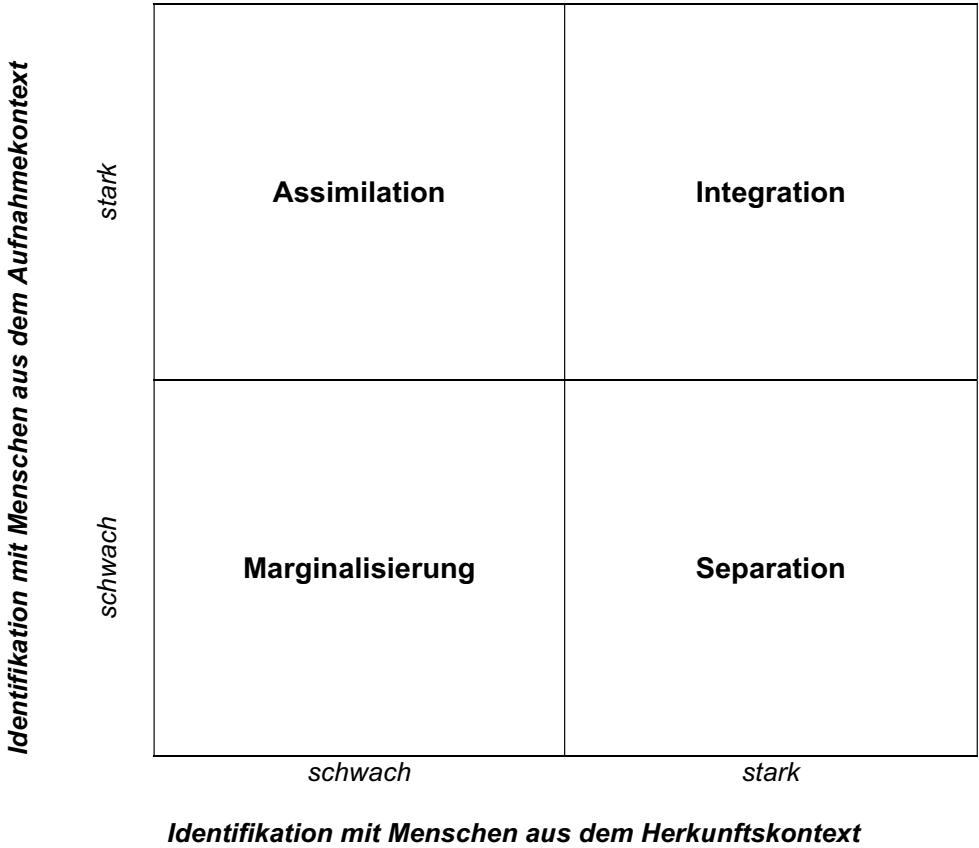


Abb. 1: Kulturelle Identitätsorientierungen (in Anlehnung an Berry, 1980, 1997)

Empirisch wurde die kulturelle Identität von Jugendlichen aus zugewanderten Familien bisher meist in klassischen Einwanderungsländern wie Kanada oder den USA untersucht. Es ist jedoch anzunehmen, dass die unterschiedlichen politischen Rahmenbedingungen und gesellschaftlichen Einstellungen gegenüber Zuwanderung in verschiedenen Aufnahmekontexten maßgeblich mitbestimmen, welche Akkulturationsstrategien bzw. Identitätsorientierungen Menschen mit Zuwanderungshintergrund entwickeln und wie adaptiv diese Strategien im jeweiligen Kontext sind (Berry, 1997; Bourhis et al., 1997; Phinney et al., 2001). Daher ist fraglich, ob die Befunde aus klassischen Einwanderungsländern auf den deutschen Kontext mit seiner vergleichsweise jüngeren Entwicklung hin zum Einwanderungsland übertragbar sind.

Weiterhin wird diskutiert, dass für die Entwicklung der kulturellen Identität von Heranwachsenden mit Zuwanderungshintergrund neben dem Aufnahmekontext auch Hintergrundmerkmale der jeweiligen Zuwanderungsgruppe, wie etwa ihr spezifischer Herkunftskontext, eine wichtige Rolle spielen können (z.B. Phinney et al., 2001). Daher ist es sinnvoll, in Studien zur kulturellen Identität die verschiedenen Herkunftsgrup-

pen differenziert zu betrachten (vgl. z.B. Fuligni, Witkow & Garcia, 2005; Horenczyk, 2010; Rumbaut, 1994). Darüber hinaus gibt es Hinweise darauf, dass Mädchen häufiger beide Kulturen integrieren (Berry et al., 2006b) und sich selbst als beiden Kulturen zugehörig beschreiben als ihre männlichen Peers (z.B. Rumbaut, 1994).

Zudem wird oft angenommen, dass die Aufenthaltsdauer der Einwandererfamilien im Aufnahmeland mit der kulturellen Identität zusammenhängt. Demnach unterscheiden sich Einwanderer im Laufe der Generationen zunehmend weniger von den Mitgliedern der Aufnahmegesellschaft (z.B. Alba & Nee, 1997, 2003; Berry et al., 2006). Entsprechend ist davon auszugehen, dass sich die Identifikation mit der Herkunftsgruppe mit zunehmender Aufenthaltsdauer reduziert, die Identifikation mit dem Aufnahmeland hingegen zunimmt. Dieser Befund zeigt sich relativ konsistent in empirischen Untersuchungen (Berry et al., 2006; Phinney, 1990). Auch in Familien mit nur einem im Ausland geborenen Elternteil dürfte der Herkunftskontext in der Sozialisation der Jugendlichen eine geringere, der Aufnahmekontext hingegen eine wichtigere Rolle spielen als in Familien mit zwei zugewanderten Elternteilen. Darum ist weiterhin anzunehmen, dass sich Jugendliche mit nur einem im Ausland geborenen Elternteil stärker mit dem Aufnahmekontext und schwächer mit der Herkunftsgruppe identifizieren als Jugendliche mit zwei im Ausland geborenen Elternteilen. Diese Annahmen sollen in der vorliegenden Studie für den deutschen Kontext untersucht werden.

1.2 Theoretische Annahmen zum Zusammenhang von Akkulturation und kultureller Identität mit der schulischen Adaption bei Jugendlichen mit Zuwanderungshintergrund

Eine zentrale Frage der Zuwanderungsforschung ist, welche Faktoren beeinflussen, inwieweit bei Zuwanderern und ihren Nachkommen eine erfolgreiche Adaption an den Aufnahmekontext erfolgt. Ward (1996; Ward & Kennedy, 1993) unterscheidet dabei die Aspekte der *psychologischen Adaption* und der *soziokulturellen Adaption*. Psychologische Adaption bezieht sich auf das psychische Wohlbefinden, die psychische Gesundheit, die Zufriedenheit und das allgemeine Selbstwertgefühl. Soziokulturelle Adaption umfasst dagegen soziale und kulturelle Fertigkeiten und Orientierungen, die bei der Bewältigung des Alltags im Familien-, Arbeits- oder Schulkontext im Aufnahmeland hilfreich sind. Häufig wird schulische Adaption anhand von schulbezogenen Einstellungen und Affekten oder – auf der negativen Seite – Verhaltensauffälligkeiten untersucht (Berry et al., 2006a, 2006b; Horenczyk, 2010). Auch der Erwerb von schulischen Kompetenzen kann als Merkmal der soziokulturellen Adaption aufgefasst werden, da zu einer erfolgreichen Anpassung an den schulischen Kontext nicht zuletzt die Bewältigung der dort gestellten Leistungsanforderungen gehört. Entsprechend wird im vorliegenden Beitrag schulische Kompetenz als wichtiger Aspekt der soziokulturellen Adaption betrachtet (vgl. Berry et al., 2006).

Der Zusammenhang zwischen kultureller Identität bzw. Akkulturationsorientierungen einerseits und psychischer Adaption andererseits wurde bereits mehrfach untersucht

(z.B. Berry et al., 2006a). Das Verhältnis zwischen kultureller Identität und schulischer Anpassung ist dagegen deutlich seltener analysiert worden. Zur Frage, ob bestimmte kulturelle Identitätsorientierungen besonders adaptiv für den Erwerb schulischer Kompetenzen sind, existieren unterschiedliche, teils widersprüchliche theoretische Positionen.

Position 1: (Neo-)assimilationstheoretische Positionen betonen die Adaptivität einer starken Orientierung am Aufnahmekontext für die Adaption in der Aufnahmegesellschaft (z.B. Alba & Nee, 2003; Esser, 2006, 2009). Demnach sollten Jugendliche mit assimilierter und integrierter kultureller Identität, die beide durch ein starkes Zugehörigkeitsgefühl zum Aufnahmekontext gekennzeichnet sind, besonders gute schulische Leistungen erbringen. Den auf die eigene Herkunftsgruppe bezogenen Ressourcen wird von der Assimilationstheorie hingegen keine Bedeutung beigemessen. Für ein zusätzlich stark ausgeprägtes Zugehörigkeitsgefühl zum Herkunftskontext lässt sich aus dieser Position also kein Vorteil ableiten. Integrierte Jugendliche sollten demnach keine höheren schulischen Kompetenzen erwerben als Jugendliche mit assimilierter kultureller Identität. Ob für eine starke Zugehörigkeit zum Herkunftskontext neutrale oder nachteilige Effekte zu erwarten sind, geht hingegen nicht eindeutig aus der Theorie hervor.

Position 2: Im Gegensatz dazu gehen eher pluralistisch orientierte Positionen davon aus, dass eine starke Orientierung am Herkunftskontext – die sowohl bei einer integrieren als auch bei einer separierten Identitätsorientierung vorliegt – positiv mit der Anpassung an die Aufnahmegesellschaft zusammenhängt. Es wird angenommen, dass eine starke Einbindung in die Herkunftsgruppe Ressourcen mobilisiert, die eine Adaption an die Aufnahmegesellschaft unterstützt (z.B. Phinney, 1990; Portes & Rumbaut, 2001; Zhou, 1997). So wird zum Beispiel argumentiert, dass eine starke Identifikation mit der Herkunftsgruppe Jugendliche aus zugewanderten Familien vor negativen Auswirkungen von Diskriminierungserfahrungen und Vorurteilen schütze (z.B. Caldwell, Kohn-Wood, Schmeelk-Cone, Chavous & Zimmerman, 2004; Wong, Eccles & Sameroff, 2003). Oft wird von diesen Ansätzen jedoch ausschließlich die Orientierung am Herkunftskontext in den Blick genommen, während über die Rolle der Orientierung am Aufnahmekontext keine Aussagen getroffen werden. Daher ist es nicht möglich, Aussagen darüber abzuleiten, welche Zusammenhänge mit schulischen Kompetenzen für Integration im Vergleich zu Separation zu erwarten sind.

Position 3: Häufig wird darüber hinaus die Position vertreten, dass für eine gleichzeitig starke Orientierung am Herkunftskontext und am Aufnahmekontext eine besonders gute Adaption zu erwarten ist (z.B. Altschul, Oysermann & Bybee, 2008; Berry et al., 2006b; Berry, 1997). Dies wird beispielsweise damit begründet, dass eine Orientierung an beiden Kontexten der wirksamste Schutz vor schädlichen Einflüssen, wie beispielweisen Vorurteilen von Mitgliedern der Aufnahmegesellschaft gegen die eigene ethnische Gruppe, sei und daher mit der günstigsten schulischen Anpassung einhergehen sollte (Oysermann, Kimmelmeier, Fryberg, Brosh & Hart-Johnson, 2003). Diese Position legt also nahe, dass – anders als von der (Neo-)Assimilationstheorie postuliert – Jugendliche mit integrierter Identitätsorientierung das höchste Adaptivitätspo-

tenzial aufweisen und deshalb die höchsten schulischen Kompetenzen erreichen sollten, die das von assimilierten Jugendlichen erreichte Niveau noch übertrifft.

Position 4: Darüber hinaus werden oft negative Effekte der Separation, also einer starken Orientierung ausschließlich an der Herkunftsgruppe, auf die Adaption angenommen (z.B. Esser, 2009). Oysermann et al. (2003) argumentieren beispielsweise, dass Jugendliche aus zugewanderten Familien, die sich kaum als Teil der Majoritätsgesellschaft sehen und die schulischen Erfolg nicht als kongruent mit der Zugehörigkeit zu ihrer Herkunftsgruppe, sondern als typisch für die Aufnahmegesellschaft betrachten, schulisch weniger engagiert seien und daher schwächere Leistungen erzielen sollten.

Position 5: Für Jugendliche, die sich weder mit dem Aufnahme- noch mit dem Herkunftskontext stark identifizieren (Marginalisierung), wird allgemein eine besonders geringe Adaption angenommen (z.B. Berry et al., 2006; Berry, 1997). Als Ursache der schlechten Anpassung dieser Jugendlichen wird einerseits der Verlust der Herkunftskultur (*culture shedding*) angeführt, mit dem fehlende Unterstützung durch die Herkunftsgruppe einhergehe, und andererseits die als ablehnend und wenig unterstützend wahrgenommene Umgebung im Aufnahmekontext (Berry, 1997).

Die Wirkrichtung des Zusammenhangs zwischen kultureller Identität und schulischer Anpassung wird in der Literatur meist nicht explizit spezifiziert (z.B. Vedder, van de Vijver & Liebkind, 2006). Allerdings scheint in der Regel von einem Einfluss der kulturellen Identität auf die Adaption ausgegangen zu werden (z.B. Oysermann et al., 2003). Aber auch umgekehrte Einflüsse der Adaption auf die Ausprägung der kulturellen Identität sind denkbar, etwa wenn geringer schulischer Erfolg zur Ablehnung des Aufnahmekontextes führt. Insgesamt erscheint es daher plausibel, von einer wechselseitigen Beeinflussung der kulturellen Identität und Adaption auszugehen.

1.3 *Befundlage zum Zusammenhang zwischen kultureller Identität und schulischer Adaption von Jugendlichen mit Zuwanderungshintergrund*

Empirisch ist der Zusammenhang zwischen kultureller Identität und schulischem Kompetenzerwerb bisher kaum untersucht worden und die Befundlage ist äußerst heterogen. Die Ergebnisse einiger Studien unterstützen die assimilationstheoretische Perspektive, dass insbesondere die Orientierung am Aufnahmekontext zentral für eine erfolgreiche schulische Anpassung ist, während der Orientierung am Herkunftskontext wenig Bedeutung zukommt (z.B. Horenczyk, 2010; Trickett & Birman, 2005). Auch in einer der wenigen Studien, in der nicht nur selbstberichtete Noten, sondern auch Kompetenzdaten erhoben wurden und die zudem in Deutschland durchgeführt worden ist, ergaben sich keine Hinweise auf Leistungsvorteile für Jugendliche, die den Aufnahme- und den Herkunftskontext in ihr schulisches Selbstbild integrieren, im Vergleich zu Heranwachsenden, die sich im schulischen Kontext ausschließlich als Deutsche sehen (Hannover et al., 2013).

In mehreren, meist in den USA durchgeführten Studien, die allerdings ausschließlich die Identifikation mit der Herkunftsgruppe, nicht jedoch die Identifikation mit dem

Aufnahmekontext erfasst haben, zeigten sich hingegen positive Zusammenhänge zwischen einer stark ausgeprägten Orientierung an der Herkunftsgruppe bzw. der Bedeutung, die dieser Gruppe zugeschrieben wird, und Aspekten der sozio-kulturellen Anpassung an den Kontext Schule, wie z.B. akademischer Selbstwirksamkeitsüberzeugung (Oysermann, Harrison & Bybee, 2001), positiven Einstellungen gegenüber der Schule (Chavous et al., 2003; Fuligni et al., 2005) und dem meist anhand von Noten erfassten Schulerfolg (z.B. Altschul, Oysermann & Bybee, 2006; Chavous et al., 2003; Wong et al., 2003).

Einige Untersuchungen, die neben der Orientierung am Herkunftskontext auch die Identifikation mit dem Aufnahmekontext erfasst haben, identifizierten die günstigste schulische Adaption (Berry et al., 2006) bzw. die höchsten schulischen Kompetenzen (Oysermann et al., 2003) bei Jugendlichen, die sich stark an beiden Kontexten orientieren. Berry (1997) interpretiert den Forschungsstand sogar dahingehend, dass sich Integration allgemein als die günstigste Akkulturationsstrategie erwiesen habe.

Darüber hinaus ergaben verschiedene Studien, dass separierte Jugendliche schlechtere Noten erreichen als Gleichaltrige, die sich an beiden Kontexten orientieren (Oysermann et al., 2003; Altschul et al., 2008). Auch Hannover et al. (2013) berichten schwächere Schulleistungen bei Jugendlichen mit einem ausschließlich am Herkunftskontext orientierten schulischen Selbstbild im Vergleich zu Jugendlichen, die sich (auch) am Aufnahmekontext orientieren. Zudem fanden z.B. Berry et al. (2006b) die ungünstigste schulische Adaption bei Jugendlichen mit marginalisiertem Integrationsprofil.

Die Heterogenität der Befunde könnte unter anderem darauf zurückzuführen sein, dass die zitierten Studien in unterschiedlichen Aufnahmeländern durchgeführt worden sind. Denn es ist davon auszugehen, dass der Zusammenhang zwischen schulischer Adaption und kultureller Identität vom Aufnahmekontext beeinflusst wird (Bourhis et al., 1997). Beispielsweise sollte in Aufnahmeländern, die einen hohen Anpassungsdruck auf Zuwanderer ausüben, Assimilation besonders adaptiv sein, während in Kontexten mit einer unterstützenden Herkunftsgruppe eine starke Orientierung an dieser Gruppe vorteilhaft sein dürfte (vgl. Phinney et al., 2001; Portes & Zhou, 1993). Für den deutschen Kontext existieren jedoch, abgesehen von der beschriebenen Studie von Hannover et al. (2013), kaum methodisch adäquate Analysen dieses Zusammenhangs. Insbesondere fehlen Studien mit repräsentativen Stichproben, die auf die Gesamtpopulation generalisierbare Aussagen erlauben. Darüber hinaus haben nur wenige Untersuchungen, wiederum mit Ausnahme der Studie von Hannover et al. (2013), objektive Maße der schulischen Kompetenz verwendet, sondern Schulerfolg meist anhand von selbstberichteten Noten erfasst.

Die vorliegende Studie soll dazu beitragen, die beschriebenen Forschungslücken zu schließen, indem sie den Zusammenhang zwischen kultureller Identität und Schulleistung anhand von objektiven Kompetenzmessungen untersucht. Als Indikator für die Schulleistung wird die Lesekompetenz herangezogen, die sowohl als Schlüssel für den Erwerb anderer schulischer Kompetenzen als auch für die spätere Teilhabe am wirtschaftlichen und gesellschaftlichen Leben angesehen werden kann (Naumann, Artelt,

Schneider & Stanat, 2010). Entsprechend steht die Lesekompetenz in engem Zusammenhang zu anderen Indikatoren schulischer Kompetenz, wie etwa der Mathematikleistung (z.B. Klieme, Neubrand & Lüdtke, 2001). Da wiederholt deutlich wurde, dass sich verschiedene Herkunftsgruppen in ihren Anpassungsmustern unterscheiden (z.B. Rumbaut, 1994), wird in den Analysen nach dem Herkunftsland der Familien der Jugendlichen differenziert.

1.4 Ziele der Untersuchung und Hypothesen

Der vorliegende Artikel verfolgt zwei Hauptziele. Das erste Ziel besteht darin, die Muster der kulturellen Identität von Jugendlichen mit Zuwanderungshintergrund in Deutschland zu beschreiben. Dazu kann erstmalig für den deutschsprachigen Raum ein repräsentativer Datensatz herangezogen werden. Um zu prüfen, ob die Befunde über verschiedene Herkunftsgruppen, Zuwanderergenerationen und die Geschlechter generalisierbar sind, werden die Ergebnisse entsprechend differenziert. Aus den beschriebenen theoretischen Annahmen und Befunden zur Aufenthaltsdauer bzw. zum Generationsstatus und zu Geschlechterunterschieden lassen sich folgende Hypothesen ableiten:

Hypothese 1a. Die zweite Zuwanderungsgeneration verfügt häufiger über eine assimilierte und seltener über eine separierte Identitätsorientierung als die erste Generation.

Hypothese 1b. Jugendliche mit nur einem im Ausland geborenen Elternteil weisen häufiger eine assimilierte und seltener eine separierte Identitätsorientierung auf als Jugendliche mit zwei im Ausland geborenen Elternteilen.

Hypothese 2. Mädchen verfügen häufiger über eine integrierte Identitätsorientierung als Jungen.

Das zweite Ziel dieses Beitrags besteht darin, zu klären, wie die Lesekompetenz von Jugendlichen aus zugewanderten Familien in Deutschland mit ihrer kulturellen Identität zusammenhängt. Dabei wird der Einfluss verschiedener Hintergrundmerkmale berücksichtigt, von denen bekannt ist, dass sie eng mit der Lesekompetenz zusammenhängen, um den von diesen Faktoren bereinigten Zusammenhang zwischen kultureller Identität und Leseleistung zu schätzen.

Sowohl aus (neo-)assimilationstheoretischen als auch aus pluralistischen Ansätzen kann folgende Annahme für integrierte Jugendliche abgeleitet werden:

Hypothese 3. Jugendliche mit integrierter Identitätsorientierung weisen ein ähnliches Lesekompetenzniveau auf wie im Hinblick auf zentrale Merkmale des familiären Hintergrundes vergleichbare Jugendliche ohne Zuwanderungshintergrund.

Aus der (Neo-)Assimilationstheorie lässt sich zudem die folgende Hypothese für Jugendliche mit assimilierter Identitätsorientierung ableiten:

Hypothese 4. Assimilierte Jugendliche verfügen über ähnlich gute Lesekompetenzen wie Jugendliche ohne Zuwanderungshintergrund mit vergleichbaren Merkmalen des familiären Hintergrundes.

Außerdem kann auf Basis der beschriebenen theoretischen Annahmen und früherer Befunde davon ausgegangen werden, dass Jugendliche mit separierter und insbesondere marginalisierter Identitätsorientierung Schwierigkeiten haben, sich an den Kontext Schule zu adaptieren:

Hypothese 5. Jugendliche mit separierter Identitätsorientierung weisen schwächere Lesekompetenzen auf als im Hinblick auf zentrale Merkmale des familiären Hintergrundes vergleichbare Jugendliche ohne Zuwanderungshintergrund.

Hypothese 6. Jugendliche mit marginalisierter Identitätsorientierung verfügen über schwächere Lesekompetenzen als Jugendliche ohne Zuwanderungshintergrund mit vergleichbaren Merkmalen des familiären Hintergrundes.

2. Methode

Die vorliegende Untersuchung beruht auf den Daten der deutschen Ergänzungsstichprobe der PISA-Erhebung, die im Frühjahr 2009 durchgeführt worden ist. Dabei wurden neben Kompetenzen in den Bereichen Lesen, Mathematik und Naturwissenschaften auch zentrale Hintergrundmerkmale der Jugendlichen erfasst.

2.1 Definition des Zuwanderungshintergrundes

Anhand der Geburtsländer der Jugendlichen, ihrer Eltern und Großeltern lassen sich verschiedene Zuwanderergenerationen unterscheiden (Generationsstatus). Der Schülerfragebogen erfasste die Geburtsländer der Jugendlichen selbst und ihrer Eltern mit einer Länderliste sowie einer offenen Antwortkategorie für nicht in der Liste enthaltene Länder. Um die dritte Generation identifizieren zu können, fragte der Schülerfragebogen danach, in welchem Land die Eltern der Mutter und die Eltern des Vaters geboren sind (Antwortkategorien: „in Deutschland“, „in einem anderen Land“). Ergänzend wurde mit einer Länderliste im Elternfragebogen erfasst, in welchen Ländern die Großeltern des Kindes geboren sind, um anhand dieser Angaben auch die dritte Generation einzelnen Herkunftsländern zuordnen zu können. Da für die Jugendlichen der dritten Generation keine Identitätsorientierungen erhoben worden sind, wurde diese Gruppe zwar in die Stichprobenbeschreibung einbezogen, aus den Zusammenhangsanalysen zur kulturellen Identität jedoch ausgeschlossen.

Der Generationsstatus der Jugendlichen wird wie folgt definiert:

Ohne Zuwanderungshintergrund:	Beide Elternteile und mindestens drei Großeltern in Deutschland geboren.
Ein Elternteil im Ausland geboren:	Ein Elternteil im Ausland, ein Elternteil in Deutschland geboren.
Erste Generation:	Beide Elternteile und die bzw. der Jugendliche im Ausland geboren.
Zweite Generation:	Beide Elternteile im Ausland geboren, die bzw. der Jugendliche in Deutschland geboren.
Dritte Generation:	Mindestens zwei Großeltern im Ausland geboren, beide Elternteile in Deutschland geboren.

Zusätzlich wird anhand der Geburtsländer zwischen verschiedenen Herkunftsgruppen unterschieden. Die drei größten Herkunftsgruppen werden in den Analysen separat ausgewiesen (Gebiet der ehem. UdSSR, Türkei, Polen). Wenn mindestens ein Elternteil bzw. mindestens zwei Großeltern im Ausland geboren sind, werden die Jugendlichen dem Herkunftsland der zugewanderten Familienmitglieder zugeordnet. Sind die Eltern bzw. Großeltern aus unterschiedlichen Ländern zugewandert, wurden diese Jugendlichen aus den Analysen ausgeschlossen, um eine eindeutige Interpretation der Ergebnisse zu gewährleisten.

2.2 Stichprobe

Die Untersuchung wurde in 393 neunten Klassen an 202 Schulen durchgeführt. Im Durchschnitt waren die Jugendlichen zum Untersuchungszeitpunkt 15,6 Jahre alt (13 bis 19 Jahre). In die Ergänzungsstichprobe wurden Gymnasien, Real- und Hauptschulen, integrierte Gesamtschulen und Schulen mit mehreren Bildungsgängen einbezogen. Die Stichprobe umfasst insgesamt 9461 Schülerinnen und Schüler, wobei für fast jeden zehnten Jugendlichen ($N = 837$) die Angaben zum Zuwanderungshintergrund fehlen. Von den verbleibenden 8624 Jugendlichen haben 29,2% einen Zuwanderungshintergrund.² Die Verteilung der Schülerinnen und Schüler nach Generationsstatus und Herkunftsländern ist in Tabelle 1 dargestellt.

2 Dieser Anteil fällt signifikant ($p < .05$) höher aus als im nationalen Bericht über PISA 2009 (25,6%, vgl. Stanat, Rauch & Segeritz, 2010). Dies ist im Wesentlichen darauf zurückzuführen, dass in der vorliegenden Stichprobe auch Jugendlichen der dritten Generation ein Zuwanderungsstatus zugewiesen wurde.

	% (gewichtet)	SE	N (absolut)
nicht zuzuordnen	9.4	0.5	837
zuzuordnen	90.6	0.5	8624
	gültige % (gewichtet)	SE	N (absolut)
ohne Zuwanderungshintergrund	70.8	1.5	6146
mit Zuwanderungshintergrund	29.2	1.5	2478
Generationsstatus			
ein Elternteil im Ausland geboren	8.7	0.4	751
erste Generation	5.4	0.4	446
zweite Generation	11.9	1.0	992
dritte Generation	3.2	0.2	289
Herkunftsländer			
ehem. UdSSR	5.1	0.5	459
Türkei	6.5	0.6	521
Polen	3.3	0.3	280
anderes Herkunftsland	14.2	0.7	1218

Anmerkung: Gültige % beziehen sich auf Schülerinnen und Schüler, denen eindeutig ein Zuwanderungsstatus zugeordnet werden konnte.

Tab. 1: Prozentuale Anteile von Schülerinnen und Schülern der neunten Jahrgangsstufe in Deutschland nach Merkmalen des Zuwanderungshintergrunds

Die Herkunftsgruppen setzen sich in unterschiedlichen Anteilen aus den verschiedenen Zuwanderergenerationen zusammen (s. Tabelle 2). So gehört der überwiegende Anteil der Jugendlichen, deren Familie aus der ehemaligen UdSSR stammt, zur ersten Generation (51%), wohingegen in den Gruppen der Jugendlichen türkischer und polnischer Herkunft der jeweils größte Anteil der zweiten Generation zuzuordnen ist (70% bzw. 41%). Von den Jugendlichen aus anderen Herkunftsländern hat der größte Anteil nur ein im Ausland geborenes Elternteil (42%).

	gültige % (gewichtet)	SE	N (absolut)
ehem. UdSSR			
ein Elternteil im Ausland geboren	8.9	1.6	42
erste Generation	50.8	3.8	225
zweite Generation	38.4	4.2	184
dritte Generation	1.8	0.7	8
Türkei			
ein Elternteil im Ausland geboren	21.5	1.9	113
erste Generation	7.6	1.5	39
zweite Generation	70.4	2.2	366
dritte Generation	0.5	0.3	3
Polen			
ein Elternteil im Ausland geboren	24.7	2.6	73
erste Generation	13.5	2.5	32
zweite Generation	41.4	3.2	115
dritte Generation	20.3	3.2	60
anderes Herkunftsland			
ein Elternteil im Ausland geboren	42.1	1.8	523
erste Generation	13.3	1.2	150
zweite Generation	27.5	1.9	327
dritte Generation	17.1	1.3	218

Anmerkung: Gültige % beziehen sich auf Schülerinnen und Schüler, denen eindeutig ein Zuwanderungsstatus zugeordnet werden konnte.

Tab. 2: Prozentuale Anteile von Schülerinnen und Schülern mit Zuwanderungshintergrund der neunten Jahrgangsstufe in Deutschland nach Herkunftsland und Generationsstatus

2.3 Messinstrumente

Kulturelle Identitätsorientierung

Die kulturelle Identität der Jugendlichen wurde zweidimensional anhand ihres Zugehörigkeitsgefühls zu den Menschen des Aufnahmekontextes (Deutschland) und zu den Menschen des Herkunftskontextes (z.B. Türkei) erfasst. Dazu sollten sie zum einen zu

„a) Den Menschen aus dem Land deiner Eltern“ und zum anderen zu „b) Den Menschen aus Deutschland“ die folgende Frage beantworten: „Wie sehr fühlst du dich den folgenden Herkunftsgruppen zugehörig?“. Die Antwortskala war dreistufig (überhaupt nicht – etwas – sehr). Zur Bildung der vier kulturellen Identitätsorientierungen wurde, in Anlehnung an Berry (1980, 1997), die dreistufige Skala dichotomisiert und bestimmt, ob sich die Jugendlichen der jeweiligen Gruppe eher stark („sehr“) oder eher schwach („überhaupt nicht“ oder „etwas“) zugehörig fühlen. Anschließend wurden die Angaben kombiniert, um die Identitätsorientierungen Assimilation, Integration, Marginalisierung und Separation zu bilden (vgl. Abbildung 1).

Der Fragebogen filtert so, dass die kulturelle Identität nur von Jugendlichen der 1. und 2. Generation sowie von mit einem im Ausland geborenen Elternteil erhoben wurde, nicht jedoch von Heranwachsenden der 3. Generation. Daher wurde die 3. Generation nicht in den Analysen zur kulturellen Identität einbezogen. Von den 2189 Jugendlichen, von denen Angaben zur kulturellen Identität erhoben werden sollten, beantworteten 83,5% die Fragen vollständig, so dass die Identitätsorientierung von 1862 Jugendlichen bestimmt werden konnte. Während der Anteil der Ausfälle in der 1. und 2. Generation fast identisch ist (11% bzw. 12%), fehlen die Werte von überproportional vielen Jugendlichen mit nur einem im Ausland geborenen Elternteil (27%).³ Die höheren Ausfälle in dieser Gruppe sind vermutlich darauf zurückzuführen, dass diese Jugendlichen den Filter im Fragebogen („Bitte beantworte die folgenden Fragen nur, wenn deine Eltern nicht in Deutschland geboren sind. Wenn deine Mutter und dein Vater in Deutschland geboren sind, dann mache bitte weiter mit Frage X“) häufig so interpretierten, dass die folgenden Fragen nur dann beantwortet werden sollten, wenn beide Eltern im Ausland geboren sind.

Um Hinweise auf die Validität der Identitätsorientierungen zu erhalten, wurde ihr Zusammenhang mit einem weiteren Indikator der kulturellen Identität analysiert, nämlich den Antworten auf die Frage: „Welcher dieser Herkunftsgruppen fühlst du dich am stärksten zugehörig?“. Hier konnte als Antwort entweder „Den Menschen aus dem Land meiner Eltern“ oder „Den Menschen aus Deutschland“ gewählt werden. Erwartungsgemäß fühlen sich über 92% der separierten Jugendlichen am stärksten den Menschen aus dem Land ihrer Eltern zugehörig und nur 4% den Menschen aus Deutschland. Umgekehrt fühlen sich 90% der assimilierten Jugendlichen am stärksten den Menschen aus Deutschland zugehörig und nur 8% den Menschen aus dem Land ihrer Eltern. Integrierte und marginalisierte Schülerinnen und Schüler entschieden sich zu etwa gleichen Teilen für die Menschen aus Deutschland (45% bzw. 43%) und für die Menschen aus dem Land ihrer Eltern (45% bzw. 39%).

3 Überzufällig viele Jugendliche mit fehlenden Werten stammen zudem aus Familien, in denen Deutsch gesprochen wird und die aus „anderen Herkunftsländern“ bzw. aus Polen zugewandert sind. Diese Merkmale kommen bei Jugendlichen mit nur einem im Ausland geborenen Elternteil besonders häufig vor.

Lesekompetenz

Als Indikator für die schulische Leistung wird die Lesekompetenz der Schülerinnen und Schüler herangezogen. Diese wurde mit dem PISA-Test erfasst, der sich am Literacy-Konzept orientiert (vgl. z.B. Naumann et al., 2010), und nach Modellen der Item-Response-Theorie skaliert (vgl. z.B. van der Linden & Hambleton, 1997). Bei den Werten, die in den Analysen verwendet werden, handelt es sich um fünf *Plausible Values*, die mittels Marginal-Maximum-Likelihood-Schätzverfahren und latenter Regression erzeugt wurde (vgl. Prenzel, Carstensen, Frey, Drechsel & Rönnebeck, 2007). In der Analysestichprobe der vorliegenden Studie hat die Lesekompetenzskala einen Mittelwert von 494 und eine Standardabweichung von 87 Punkten.

Kontrollvariablen

Zudem wurden wichtige Merkmale, die eng mit schulischer Kompetenz zusammenhängen (z.B. Stanat et al., 2010; Prenzel et al., 2007), als Kontrollvariablen in die Analysen einbezogen. Dabei handelt es sich um das Geschlecht der Jugendlichen, den sozio-ökonomischen Status der Familie, das Bildungsniveau der Eltern, die in der Familie vorhandenen kulturellen Ressourcen (Kulturgüter) sowie die in der Familie gesprochene Sprache (Familiensprache). Als Indikator der sozio-ökonomischen Stellung der Familie wurde der Index für den höchsten beruflichen Status der Mutter bzw. des Vaters (HISEI) gebildet (vgl. OECD, 2009). Fehlende Werte im HISEI wurden mit einem erweiterten Hintergrundmodell für die nationale Ergänzungsstichprobe durch ein Maximum-Likelihood-Verfahren imputiert (vgl. Carstensen, Frey, Walter & Knoll, 2007; Prenzel et al., 2007; OECD, 2009). Um das Bildungsniveau der Eltern als kontinuierliche Variable abzubilden, wurden die Schülerangaben zu den höchsten erreichten Bildungsabschlüssen der Eltern in die Anzahl der Ausbildungsjahre umgerechnet. Der Index für die Kulturgüter beruht auf Schülerangaben zum Vorhandensein von klassischer Literatur, Gedichtbänden und Kunstwerken im Elternhaus (vgl. OECD, 2011).

Familiensprache

Als weiteres wichtiges Hintergrundmerkmal wurde die in der Familie gesprochene Sprache in die Analysen einbezogen, die sowohl mit der Lesekompetenz (z.B. Müller & Stanat, 2006) als auch mit der kulturellen Identität (Vedder, Horenczyk & Liebkind, 2006) zusammenhängt. Es ist anzunehmen, dass der Zusammenhang zwischen kultureller Identität und Lesekompetenz zumindest teilweise über die Familiensprache vermittelt wird. Deshalb handelt es sich bei den Analysen, in denen die in der Familie gesprochene Sprache kontrolliert wird, um eine sehr konservative Prüfung des Zusammenhangs zwischen kultureller Identität und Lesekompetenz.

Zur Erfassung des familiären Sprachgebrauchs wurden die Schülerinnen und Schüler gefragt: „Welche Sprache sprichst du zuhause am häufigsten?“. Dieser Frage folgte eine Sprachliste. In den Analysen wird zwischen den Kategorien „Deutsch“ (0) und „eine andere Sprache“ (1) unterschieden.

3. Ergebnisse

3.1 *Kulturelle Identität von Jugendlichen mit Zuwanderungshintergrund in Deutschland*

Etwa ein Viertel der Neuntklässler aus zugewanderten Familien weist eine integrierte Identitätsorientierung auf, identifiziert sich also sowohl stark mit dem Aufnahmekontext Deutschland als auch mit dem Herkunftskontext (s. Tabelle 3). Ein weiteres Viertel fühlt sich dem Aufnahmekontext stark, dem Herkunftskontext hingegen nicht oder nur etwas zugehörig (assimiliert). Insgesamt identifiziert sich also jeder zweite Jugendliche mit Zuwanderungshintergrund stark mit dem Aufnahmekontext. Dagegen identifizieren sich 28% der Jugendlichen stark mit dem Herkunftskontext, aber nur schwach mit dem Aufnahmekontext (separiert). Eine marginalisierte Identitätsorientierung, also eine schwache Identifikation sowohl mit dem Herkunftskontext als auch mit dem Aufnahmekontext Deutschland, kommt mit 20% am seltensten vor.

Entgegen den Erwartungen ist der Anteil assimilierter und separierter Jugendlicher in der ersten und zweiten Zuwanderungsgeneration nahezu identisch (Hypothese 1a). Allerdings finden sich erwartungsgemäß besonders viele assimilierte und besonders wenige separierte Jugendliche in der Gruppe der Heranwachsenden mit nur einem im Ausland geborenen Elternteil (Hypothese 1b). Der Anteil integrierter und marginalisierter Jugendlicher ist in allen drei Generationsgruppen etwa gleich groß.

Ein Vergleich der kulturellen Identität von Jugendlichen aus unterschiedlichen Herkunftsländern zeigt, dass sich die Anteile integrierter, separierter und marginalisierter Jugendlicher nicht bedeutsam unterscheiden. Türkischstämmige Jugendliche sind jedoch signifikant seltener assimiliert als ihre Peers aus anderen Herkunftsländern. Zudem scheinen türkischstämmige Jugendliche tendenziell häufiger separiert zu sein als Jugendliche aus anderen Herkunftsgruppen (32% vs. 26%), dieser Unterschied kann jedoch bei einem nach Bonferroni korrigierten Signifikanzniveau von 5% nicht gegen den Zufall abgesichert werden. Außerdem zeigt sich, dass Mädchen erwartungsgemäß häufiger über eine integrierte kulturelle Identität verfügen, während Jungen häufiger marginalisiert sind (Hypothese 2).

Zudem sind enge Zusammenhänge zwischen der kulturellen Identität und der Familiensprache zu erkennen. Assimilierte Jugendliche sprechen häufiger Deutsch als eine andere Sprache zuhause, während separierte Jugendliche häufiger eine andere Sprache, also wohl vornehmlich ihre Herkunftssprache, verwenden. In den Familien integrierter und marginalisierter Heranwachsender werden Deutsch und andere Sprachen hingegen etwa gleich häufig verwendet.

	integriert		assimiliert		separiert		marginalisiert	
	gültige %	SE	gültige %	SE	gültige %	SE	gültige %	SE
mit Zuwanderungshintergrund gesamt	26	1.2	26	1.3	28	1.3	20 ^a	1.2
Generationsstatus								
ein Elternteil im Ausland geboren	26	1.9	38 ^b	2.1	19 ^b	1.4	18	1.7
erste Generation	27	2.2	19	2.2	33	2.5	21	2.8
zweite Generation	26	1.5	22	1.7	31	2.0	20	1.9
Herkunftsländer								
ehem. UdSSR	20	2.4	32	2.8	26	3.1	22	2.2
Türkei	29	3.1	16 ^c	2.4	32	2.9	24	2.6
Polen	25	2.8	28	3.4	26	3.3	20	2.3
anderes Herkunftsland	28	1.4	29	1.7	26	1.7	16	1.4
Geschlecht								
Mädchen	30	1.4	28	1.6	27	1.7	15	1.3
Jungen	22 ^d	1.8	25	1.6	29	1.8	25 ^d	1.7
Familiensprache								
Deutsch	25	1.7	42	1.9	14	1.3	19	1.3
andere Sprache	26	2.0	12 ^e	1.9	44 ^e	2.5	18	1.9

Bei Mehrfachvergleichen wurde das Signifikanzniveau nach Bonferroni korrigiert, so dass alle als signifikant berichteten Unterschiede auch nach der Adjustierung signifikant sind ($p < .05$).

^a signifikanter Unterschied zu allen drei anderen Identitätsorientierungen

^b signifikanter Unterschied zu den beiden anderen Generationsstatusgruppen

^c signifikanter Unterschied zu allen anderen Herkunftgruppen

^d signifikanter Geschlechterunterschied

^e signifikanter Unterschied zur Gruppe mit deutscher Familiensprache

Tab. 3: Prozentuale Anteile von Schülerinnen und Schülern mit Zuwanderungshintergrund nach Identitätsorientierung und Generationsstatus, Herkunftsländern, Geschlecht sowie Familiensprache

3.2 Zusammenhang zwischen kultureller Identität und Lesekompetenz

Zusammenhang zwischen kultureller Identität und Lesekompetenz über alle Herkunftsgruppen

Um das Zusammenspiel von kultureller Identitätsorientierung und Lesekompetenz und damit die Hypothesen 3-6 zu überprüfen, wurde eine Reihe von Regressionsanalysen durchgeführt (s. Tabelle 4). Dabei wurde die Lesekompetenz von Jugendlichen mit Zuwanderungshintergrund unterschiedlicher Identitätsorientierung mit der Lesekompetenz von Jugendlichen ohne Zuwanderungshintergrund verglichen. Nur diejenigen Neuntklässler, für die vollständige Daten auf allen berücksichtigten Variablen vorlagen, wurden in die Analysen einbezogen ($N = 7023$).

Modell 1, in das lediglich die Identitätsorientierung als unabhängige Variable eingeht, zeigt einen signifikanten Zusammenhang zwischen der Identitätsorientierung und der Lesekompetenz. Sowohl integrierte als auch separierte und marginalisierte Jugendliche erreichen eine niedrigere Lesekompetenz als Schülerinnen und Schüler ohne Zuwanderungshintergrund, wobei der Nachteil der Integrierten (–24 Punkte) knapp einer viertel Standardabweichung entspricht und der Nachteil sowohl der Separierten (–53 Punkte) als auch der Marginalisierten (–57 Punkte) über eine halbe Standardabweichung beträgt.⁴ Assimilierte Jugendliche unterscheiden sich in ihrer Lesekompetenz hingegen nicht signifikant von Jugendlichen ohne Zuwanderungshintergrund (–5 Punkte).

Anschließend wurde mit Modell 2 geprüft, ob die Zusammenhänge zwischen kultureller Identität und Lesekompetenz auch dann noch bestehen bleiben, wenn das Geschlecht und Hintergrundmerkmale der Familie berücksichtigt werden. In dem Modell wurden neben dem Geschlecht der sozioökonomische Hintergrund (HISEI), die kulturellen Ressourcen und das Bildungsniveau der Eltern kontrolliert. Nach Berücksichtigung dieser Faktoren bleiben integrierte Jugendliche weiterhin etwas hinter ihren Peers ohne Zuwanderungshintergrund zurück (–12 Punkte), während sich assimilierte Jugendliche nach wie vor nicht von Gleichaltrigen ohne Zuwanderungshintergrund unterscheiden. Separierte und marginalisierte Jugendliche weisen in diesem Modell um etwa eine drittel Standardabweichung schwächere Lesekompetenzen auf als Jugendliche ohne Zuwanderungshintergrund (jeweils –33 Punkte).

Um die Rolle des Sprachgebrauchs für den Zusammenhang zwischen kultureller Identität und Lesekompetenz zu bestimmen, wurde in Modell 3 zusätzlich die Familiensprache berücksichtigt. Integrierte Jugendliche liegen in ihrer Lesekompetenz nun nicht mehr hinter ihren Peers ohne Zuwanderungshintergrund zurück. Dieser Befund stützt die Annahme, dass Integrierte im Vergleich zu Jugendlichen ohne Zuwanderungshintergrund nicht benachteiligt sind (Hypothese 3), zumindest dann nicht, wenn der familiäre Hintergrund und die Familiensprache kontrolliert werden. Assimilierte Jugendliche erreichen in Modell 3 sogar 11 Punkte mehr auf der Kompetenzskala als Jugendliche ohne

4 Der Kompetenzunterschied im Lesen zwischen praktisch gleichaltrigen Schülerinnen und Schülern aus zwei aufeinanderfolgenden Schulstufen liegt nach Schätzungen der OECD (2010) im Durchschnitt bei etwa 39 Punkten auf der PISA-Skala.

Zuwanderungshintergrund. Jugendliche aus zugewanderten Familien, die sich hinsichtlich ihres familiären Hintergrundes nicht von Jugendlichen ohne Zuwanderungshintergrund unterscheiden, hauptsächlich Deutsch zuhause sprechen und sich stark mit den Menschen aus Deutschland sowie wenig mit den Menschen aus ihrem Herkunftsland identifizieren, scheinen also besonders erfolgreich zu sein. Hypothese 4, die für assimilierte Jugendliche im Vergleich zu Schülerinnen und Schülern ohne Zuwanderungshintergrund keine Kompetenznachteile vorhersagt, wird also ebenfalls bestätigt.

Hypothese 5, die schwächere Lesekompetenzen für separierte Jugendliche im Vergleich zu Jugendlichen ohne Zuwanderungshintergrund postuliert, wird nach Berücksichtigung aller Hintergrundmerkmale in Modell 3 nicht bestätigt, da der geringfügige Nachteil von 10 Lesekompetenzpunkten nicht gegen den Zufall abgesichert werden kann. Die in der Literatur berichteten negativen Effekte einer separierten Orientierung sind möglicherweise auf die Familiensprache zurückzuführen: nachdem dieser Faktor be-

	Modell 1		Modell 2		Modell 3	
	b	SE	b	SE	b	SE
ohne Zuwanderungshintergrund	513	2.6	496	2.4	496	2.5
mit Zuwanderungshintergrund						
integriert	-24	5.7	-12	5.4	3	5.4
assimiliert	-5	6.1	5	5.7	11	5.4
separiert	-53	6.9	-33	6.1	-10	7.0
marginalisiert	-57	6.0	-33	5.4	-20	5.4
Geschlecht ¹			31	2.3	31	2.3
HISEI ²			17	1.4	17	1.4
Kulturgüter ²			13	1.4	13	1.4
Bildungsniveau der Eltern ²			11	1.2	10	1.2
Familiensprache ³					-34	5.0
N	7023		7023		7023	
R ²	0.04		0.21		0.22	

¹ Referenzgruppe: Jungen

² z-standardisiert

³ Referenzgruppe: Deutsch als Familiensprache

fett: signifikante Partialregressionskoeffizienten ($p < .05$)

Tab. 4: Schätzung von Disparitäten in der Lesekompetenz zwischen Jugendlichen mit Zuwanderungshintergrund mit unterschiedlichen Identitätsorientierungen und Jugendlichen ohne Zuwanderungshintergrund

rücksichtigt wird, bleiben separierte Jugendliche in ihrer Lesekompetenz nicht mehr signifikant hinter ihren Peers ohne Zuwanderungshintergrund zurück. Die Annahme, dass marginalisierte Jugendliche besonders benachteiligt sind (Hypothese 6), wird hingegen auch nach Berücksichtigung der Hintergrundmerkmale und der Familiensprache gestützt. Sie liegen auch im Modell 3 20 Kompetenzpunkte hinter ihren Peers ohne Zuwanderungshintergrund zurück.

Zusammenhang zwischen kultureller Identität und Lesekompetenz innerhalb von Herkunftsgruppen

Um zu prüfen, ob sich die Zusammenhänge zwischen kultureller Identität und Lesekompetenz über verschiedene Herkunftsgruppen generalisieren lassen, wurden die Analysen nochmals entsprechend differenziert durchgeführt (s. Tabelle 5). Allgemein zeigen sich viele Gemeinsamkeiten zu den Analysen, in denen nicht nach Herkunftsländern unterschieden wurde, aber auch einige Besonderheiten.

Analog zu den Befunden über alle Herkunftsgruppen hinweg zeigt sich für Schülerinnen und Schüler mit separierter und marginalisierter Identitätsorientierung vor Kontrolle der Hintergrundmerkmale (Modell 1) in allen Herkunftsgruppen ein Lesekompetenznachteil. Dies ist im Allgemeinen auch noch nach Kontrolle der familiären Hintergrundmerkmale der Fall (Modell 2), nur in den aus dem Gebiet der ehemaligen UdSSR und aus Polen stammenden Herkunftsgruppen erreicht diese Tendenz nicht mehr konsistent das Signifikanzniveau. In Modell 2 ist zudem in fast allen Herkunftsgruppen, mit Ausnahme der türkischen, für assimilierte und integrierte Jugendliche kein Nachteil gegenüber Schülerinnen und Schülern ohne Zuwanderungshintergrund zu erkennen. Wird zusätzlich die Familiensprache kontrolliert (Modell 3), so verbessert sich in allen Herkunftsgruppen die Lesekompetenz weiter. Insbesondere assimilierte Heranwachsende weisen nun meistens – analog zu den Befunden ohne Differenzierung nach Herkunftsgruppen – zumindest tendenziell ein höheres Lesekompetenzniveau auf als autochthone Gleichaltrige. Nur für Jugendliche mit türkischen Wurzeln ist dies nicht der Fall. Generalisieren lässt sich zudem der Befund, dass marginalisierte Heranwachsende selbst nach Kontrolle der Familiensprache schwächere Lesekompetenzen als Jugendliche ohne Zuwanderungshintergrund erreichen, auch wenn sich diese Tendenz nicht in allen Herkunftsgruppen gegen den Zufall absichern lässt. Zudem verringern sich in allen Herkunftsgruppen, außer in der Gruppe mit polnischem Hintergrund, die Nachteile der separierten Jugendlichen nach Kontrolle der Familiensprache so weit, dass sie nicht mehr statistisch signifikant sind.

Für die türkische Herkunftsgruppe fällt auf, dass sie durchgängig besonders stark benachteiligt ist. Dies gilt in dieser Gruppe auch für Jugendliche mit Identitätsorientierungen, bei denen in anderen Herkunftsgruppen keine Nachteile in der Lesekompetenz zu beobachten sind. So sind assimilierte und integrierte Schülerinnen und Schüler türkischer Herkunft auch nach Berücksichtigung des familiären Hintergrundes (Modell 2) noch deutlich benachteiligt, und selbst nach zusätzlicher Kontrolle der Familiensprache (Modell 3) sind tendenziell noch Nachteile für türkischstämmige Jugendliche aller Identitätsorientierungen erkennbar, die allerdings bei den marginalisierten Jugendlichen besonders

	Modell 1		Modell 2		Modell 3		N (absolut)
	b	SE	b	SE	b	SE	
ohne Zuwanderungs- hintergrund	513	2.6	496	2.5	496	2.5	5680
ehem. UdSSR							
integriert	-13	15.8	-7	13.6	11	13.1	51
assimiliert	11	7.4	17	7.0	23	7.0	97
separiert	-29	10.3	-14	9.6	6	10.0	63
marginalisiert	-47	14.0	-25	12.1	-13	11.6	53
Türkei							
integriert	-69	9.2	-40	9.7	-22	9.8	79
assimiliert	-60	16.1	-35	12.8	-22	12.4	54
separiert	-79	9.5	-42	9.1	-19	10.1	103
marginalisiert	-90	11.0	-54	11.4	-37	10.9	69
Polen							
integriert	-16	3.7	-9	12.3	2	12.1	45
assimiliert	6	15.8	14	14.9	14	15.0	49
separiert	-61	14.8	-51	14.0	-33	14.5	31
marginalisiert	-51	25.4	-24	25.9	-16	25.4	24
anderes Herkunftsland							
integriert	-4	7.9	1	7.7	10	7.9	172
assimiliert	3	8.6	9	8.0	13	7.6	213
separiert	-45	9.4	-31	8.4	-12	8.8	143
marginalisiert	-41	8.1	-25	8.2	-17	8.0	97
Geschlecht ¹			31	2.3	31	2.3	
HISEI ²			17	1.4	17	1.4	
Kulturgüter ²			13	1.4	13	1.4	
Bildungsniveau der Eltern ²			10	1.2	10	1.2	
Familiensprache ³					-30	5.1	
N	7023		7023		7023		
R ²	0.05		0.22		0.22		

¹ Referenzgruppe: Jungen² z-standardisiert³ Referenzgruppe: Deutsch als Familiensprache**fett:** signifikante Partialregressionskoeffizienten ($p < .05$)

Tab. 5: Schätzung von Disparitäten in der Lesekompetenz zwischen Jugendlichen mit Zuwanderungshintergrund mit unterschiedlichen Identitätsorientierungen nach Herkunftsland der Familie und Jugendlichen ohne Zuwanderungshintergrund

stark ausgeprägt und auch nur für diese Gruppe statistisch signifikant sind. In der polnischen Herkunftsgruppe fällt auf, dass separierte Jugendliche besonders benachteiligt sind, sogar stärker als marginalisierte Jugendliche.

4. Diskussion

In der vorliegenden Studie wurde erstmalig im deutschen Kontext auf der Grundlage von Daten einer repräsentativen Stichprobe und anhand von objektiven Kompetenzmessungen der Zusammenhang zwischen kultureller Identität und schulbezogenen Leistungen von Jugendlichen untersucht. Die Ergebnisse zeigen, dass sich mehr als die Hälfte der Jugendlichen aus zugewanderten Familien stark mit dem Aufnahmekontext Deutschland identifiziert (assimiliert und integriert) und sich ein weiteres Drittel ausschließlich ihrem Herkunftskontext stark zugehörig fühlt (separiert). Eine schwache Identifikation mit beiden Kontexten (marginalisiert) berichtet ein Fünftel der Jugendlichen.

Entgegen den Erwartungen identifiziert sich die zweite Zuwanderungsgeneration nicht stärker mit dem Aufnahmekontext und schwächer mit dem Herkunftskontext als die erste Generation. Eine längere Aufenthaltsdauer scheint also nicht zwangsläufig mit der Aufgabe der Identifikation mit dem Herkunftskontext zugunsten des Aufnahmekontextes einherzugehen. In der Gruppe der Jugendlichen mit nur einem im Ausland geborenen Elternteil finden sich erwartungsgemäß mehr assimilierte und weniger separierte Heranwachsende als bei Jugendlichen aus Familien, in denen beide Eltern im Ausland geboren sind. Für die Identität dieser Schülerinnen und Schüler scheint ihr Zuwanderungshintergrund also eine geringere Rolle zu spielen. Allerdings war in dieser Gruppe die Anzahl der fehlenden Werte erhöht, so dass unsicher ist, ob die Befunde auch für Schülerinnen und Schüler gelten, für die keine Angaben zu ihrer Identitätsorientierung vorliegen.

Der Vergleich von Jugendlichen aus verschiedenen Herkunftskontexten ergab kaum signifikante Unterschiede in der kulturellen Identitätsorientierung. Aufgrund der geringen Fallzahlen in den einzelnen Zellen ist allerdings die Teststärke dieser Analysen eingeschränkt. Es zeigte sich, dass Jugendliche, deren Familien aus der Türkei stammen, seltener assimiliert sind als Jugendliche aus anderen Herkunftskontexten. Jugendliche aus türkischstämmigen Familien zeichnen sich also offenbar dadurch aus, dass die Orientierung an ihrer Herkunftskultur selten durch die Orientierung an der Aufnahmekultur ersetzt wird. Möglicherweise ist dies darauf zurückzuführen, dass türkische Familien vornehmlich als sogenannte Gastarbeiter nach Deutschland kamen und ein dauerhafter Aufenthalt für sie nicht vorgesehen war. Dieser Umstand könnte einerseits in der Aufnahmegesellschaft eine weniger inklusive Haltung gegenüber dieser Herkunftsgruppe gefördert und andererseits zu einem verstärkten Festhalten der Zuwanderungsgruppe an ihrer türkischen Identität geführt haben. Diese Annahme wird durch Hinweise darauf gestützt, dass in Deutschland negative Vorurteile gegenüber Zuwanderern türkischer Herkunft bestehen (vgl. Schofield, 2006).

Die Analysen der Lesekompetenz über alle Herkunftsgruppen hinweg weisen darauf hin, dass Heranwachsende mit Zuwanderungshintergrund, die eine integrierte kulturelle Identität aufweisen, bei vergleichbarem familiärem Hintergrund und deutscher Familiensprache in ihren Leistungen nicht hinter autochthonen Heranwachsenden zurückbleiben. Ebenso wurde deutlich, dass bei assimilierten Jugendlichen keine Nachteile in der Lesekompetenz im Vergleich zu Jugendlichen ohne Zuwanderungshintergrund auftreten, wenn der familiäre Hintergrund der Familie kontrolliert wird. Wird zusätzlich die Familiensprache berücksichtigt, erreichen assimilierte Jugendliche sogar eine signifikant höhere Lesekompetenz als die Vergleichsgruppe der Schülerinnen und Schüler ohne Zuwanderungshintergrund.

Diese Befunde sind mit der assimilationstheoretischen Position, dass die Orientierung am Aufnahmekontext für eine erfolgreiche Adaption zentral ist, gut vereinbar. Gleichzeitig widersprechen sie der Annahme, dass eine integrative kulturelle Identität besonders adaptiv sei. Dieses Befundmuster steht also im Widerspruch zu den Ergebnissen der groß angelegten Studie von Berry et al. (2006a, 2006b), in der sich für ein integriertes Profil die günstigste Anpassung zeigte. Die Abweichung könnte zum einen darauf zurückzuführen sein, dass in den beiden Studien etwas unterschiedliche Konstrukte untersucht wurden. Berry und Kollegen analysierten erstens nicht nur die kulturelle Identität an sich, sondern ein Konglomerat von Variablen wie Akkulturationsstrategien, selbstberichtete Sprachkompetenz, soziale Kontakte und familienbezogene Werte. Zweitens untersuchten sie nicht schulische Kompetenzen, sondern andere Aspekte der soziokulturellen Anpassung, wie etwa Verhaltensauffälligkeiten und Einstellungen gegenüber der Schule.

Zum anderen könnten aber auch Eigenschaften des Aufnahmekontextes für die abweichenden Befunde verantwortlich sein. Die meisten Studien, die besonders positive Effekte der Integration identifizierten, wurden in klassischen Einwanderungsländern durchgeführt. Dort könnten integrative Akkulturationsstrategien und Identitätsorientierungen besonders günstig sein, da pluralistische Einstellungen in diesen Ländern gesellschaftlich stärker unterstützt werden und somit eine gute Passung zwischen den Erwartungen der Aufnahmegesellschaft und einer integrativen Orientierung von Menschen mit Zuwanderungshintergrund vorliegen sollte (vgl. Berry, 1997; Bourhis et al., 1997). In Deutschland, das erst seit kurzer Zeit und nur zögerlich anerkennt, ein Einwanderungsland zu sein, könnte hingegen eine bessere Passung zwischen einer assimilativen Identitätsorientierung und den gesellschaftlichen Erwartungen des Aufnahmekontextes bestehen (vgl. Hannover et al., 2013; Phinney et al., 2001).

Separierte Jugendliche zeigen sowohl vor als auch nach Berücksichtigung des familiären Hintergrundes schwächere Lesekompetenzen als Jugendliche ohne Zuwanderungshintergrund. Allerdings werden diese Nachteile nicht mehr signifikant, wenn zusätzlich die Familiensprache berücksichtigt wird. Bei separierten Heranwachsenden lässt sich also der Nachteil offenbar weitgehend auf die Familiensprache zurückführen. Demnach scheint eine separierte Identitätsorientierung, die mit einem starken Gebrauch der Herkunftssprache einhergeht, was empirisch oft der Fall ist, mit deutlichen Risiken für den Lesekompetenzerwerb verbunden zu sein. Separierte Jugendliche, die haupt-

sächlich Deutsch in der Familie sprechen, weisen jedoch – im Gegensatz zu marginalisierten Heranwachsenden – keine deutlichen Kompetenznachteile auf.

Die Lesekompetenz von Jugendlichen mit marginalisierter Identitätsorientierung bleibt hingegen erwartungsgemäß auch nach Berücksichtigung der Hintergrundmerkmale und der Familiensprache hinter der von Peers ohne Zuwanderungshintergrund zurück. Diese Gruppe scheint also für eine geringe schulische Adaption besonders gefährdet zu sein. In Anbetracht dieses Befundes scheint es angemessen, diese Gruppe als marginalisiert zu bezeichnen, anstatt eine neutralere Terminologie wie indifferent zu verwenden (vgl. Maehler, 2012).

Insgesamt deutet das Befundmuster darauf hin, dass ein starkes Zugehörigkeitsgefühl zu mindestens einer kulturellen Gruppe mit einem günstigeren schulischen Kompetenzerwerb einhergeht. Dabei scheint insbesondere eine starke Identifikation mit dem Aufnahmekontext wichtig zu sein. Jugendliche, deren Zugehörigkeitsgefühl sowohl zum Aufnahmekontext als auch zum Herkunftskontext schwach ausgeprägt ist, haben hingegen, selbst bei vergleichbarem familiärem Hintergrund und wenn in ihrer Familie Deutsch gesprochen wird, ein erhöhtes Risiko, schwache Lesekompetenzen auszubilden.

Auch die Befunde innerhalb der Herkunftsgruppen bestätigen dieses Muster weitgehend. Allerdings zeigt sich für die Gruppe türkischer Herkunft die Besonderheit, dass sie allgemein sehr stark benachteiligt ist und selbst Jugendliche mit assimilierter und integrierter Identitätsorientierung nach Kontrolle des familiären Hintergrundes in ihrer Lesekompetenz noch deutlich hinter der Vergleichsgruppe zurückbleiben. Wird zusätzlich die Familiensprache berücksichtigt, so sind marginalisierte Heranwachsende wiederum am stärksten benachteiligt, aber auch für Jugendliche mit anderen Identitätsorientierungen bleiben tendenziell noch Nachteile bestehen. Die Identitätsorientierung kann also die in der Gruppe der türkischstämmigen Schülerinnen und Schüler besonders ausgeprägten Kompetenznachteile nicht vollständig erklären.

Eine Einschränkung der vorliegenden Untersuchung besteht in der Operationalisierung der kulturellen Identität. Die Identifikation mit dem Herkunftskontext bzw. Aufnahmekontext wurde jeweils nur anhand eines einzelnen Items erfasst. Allerdings untermauert die Übereinstimmung der Antworten auf diese Items mit der ebenfalls erfassten Entscheidung in Bezug auf das stärkste Zugehörigkeitsgefühl zu einem der beiden Kontexte die Validität der Angaben. Dennoch wäre für zukünftige Studien eine breitere Operationalisierung des Konstrukts wünschenswert.

Hinsichtlich der Operationalisierung der kulturellen Identität ist zudem zu berücksichtigen, dass aufgrund des gewählten Cut-Offs eine in der vorliegenden Studie als „schwach“ kodierte Identifikation durchaus durch ein gewisses Maß an Zugehörigkeitsgefühl charakterisiert sein kann. Wenn beispielsweise von einer marginalisierten Identitätsorientierung gesprochen wird, so kann eine „etwas“ ausgeprägte Zugehörigkeit zu einem der beiden oder zu beiden Kontexten angegeben worden sein. Wären auch diejenigen Jugendlichen, die angegeben haben, sich „etwas“ mit einem Kontext zu identifizieren als „stark“ identifiziert eingestuft worden, so wären die Anteile der marginalisierten Jugendlichen, aber auch der separierten und assimilierten Heranwachsenden

geringer ausgefallen. Die vorliegende Operationalisierung wurde gewählt, um die Anschlussfähigkeit an andere Studien zum Zusammenhang von kultureller Identität und sozialer Anpassung im deutschsprachigen Raum zu gewährleisten, in denen die mittlere Kategorie ebenfalls der gering ausgeprägten Zugehörigkeit zugeordnet wurde (vgl. Esser, 2006, 2009).

Eine weitere Begrenzung der vorliegenden Studie liegt in ihrem querschnittlichen Design, das es nicht erlaubt, die zugrunde liegende Wirkrichtung des Zusammenhangs von kultureller Identität und schulischer Kompetenz zu untersuchen. Theoretisch sind beide Wirkrichtungen denkbar. Es ist beispielsweise plausibel, anzunehmen, dass sich eine starke Orientierung am Aufnahmekontext günstig auf den Schulerfolg auswirkt. Umgekehrt könnten sich Jugendliche mit hohen schulischen Kompetenzen aber auch vermehrt mit der Aufnahmegesellschaft, ihren Normen und Institutionen identifizieren. Daher scheint es naheliegend, ähnlich wie beim Zusammenspiel von Schulleistung und Selbstkonzept (z.B. Möller & Trautwein, 2009), von einer wechselseitigen Beeinflussung von kultureller Identität und schulischer Kompetenz auszugehen.

Insgesamt macht unsere Studie deutlich, dass sich die meisten in Deutschland lebenden Jugendlichen mit Zuwanderungshintergrund stark mit dem Aufnahmekontext und/oder dem Herkunftskontext identifizieren. Eine Minderheit identifiziert sich jedoch mit keinem der beiden Kontexte stark. Diese Minderheit ist besonders gefährdet, schwache schulbezogene Kompetenzen aufzuweisen. Auch Jugendliche, die sich stark mit dem Herkunftskontext, nicht aber mit dem Aufnahmekontext identifizieren und zuhause wenig Deutsch sprechen, haben ein erhöhtes Risiko, schwache schulische Kompetenzen aufzuweisen. Diese Befunde unterstreichen die Bedeutung der aktuell in der öffentlichen Debatte diskutierten Frage, wie es in Deutschland besser gelingen kann, eine Kultur zu schaffen, in der sich Zuwanderer und ihre Nachkommen willkommen und zugehörig fühlen. Ziel schulischer und allgemein gesellschaftlicher Anstrengungen sollte es sein, insbesondere marginalisierten, aber auch stark separierten Konstellationen entgegenzuwirken.

Literatur

- Alba, R., & Nee, V. (1997). Rethinking assimilation theory for a new era of immigration. *International Migration Review*, 3, 826-874.
- Alba, R., & Nee, V. (2003). *Remaking the American mainstream. Assimilation and the contemporary immigration*. Cambridge: Harvard University Press.
- Altschul, I., Oysermann, D., & Bybee, D. (2006). Racial-ethnic identity in mid-adolescence: Content and change as predictors of academic achievement. *Child Development*, 77(5), 1155-1169.
- Altschul, I., Oysermann, D., & Bybee, D. (2008). Racial-ethnic self-schemas and segmented assimilation: Identity and the academic achievement of Hispanic youth. *Social Psychology Quarterly*, 71(3), 302-320.
- Berry, J. W. (1980). Acculturation as varieties of adaptation. In A. Padilla (Hrsg.), *Acculturation: Theory, models, and some new finding* (S. 9-25). Boulder: Westview.
- Berry, J. W. (1997). Immigration, acculturation, and adaptation. *Applied Psychology: An International Review*, 46(1), 5-68.

- Berry, J. W., Phinney, J. S., Kwak, K., & Sam, D. L. (2006). Introduction: Goals and research framework for studying immigrant youth. In J. W. Berry, J. S. Phinney, D. L. Sam & P. Vedder (Hrsg.), *Immigrant youth in cultural transition: Acculturation, identity and adaptation across national contexts* (S. 1-14). Mahwah: Lawrence Erlbaum Associates Publishers.
- Berry, J. W., Phinney, J. S., Sam, D. L., & Vedder, P. (2006a). *Immigrant youth in cultural transition: Acculturation, identity, and adaptation across national contexts*. Mahwah: Lawrence Erlbaum Associates Publishers.
- Berry, J. W., Phinney, J. S., Sam, D. L., & Vedder, P. (2006b). Immigrant youth: Acculturation, identity, and adaptation. *Applied Psychology: An International Review*, 55(3), 303-332.
- Bourhis, R. Y., Moise, L. C., Perreault, S., & Senécal, S. (1997). Towards an interactive acculturation model: A social psychological approach. *International Journal of Psychology*, 32(6), 369-386.
- Caldwell, C. H., Kohn-Wood, L. P., Schmeelk-Cone, K. H., Chavous, T. M., & Zimmerman, M. A. (2004). Racial discrimination and racial identity as risk or protective factors for violent behaviors in African American young adults. *American Journal of Community Psychology*, 33(1-2), 91-105.
- Carstensen, C. H., Frey, A., Walter, O., & Knoll, S. (2007). Technische Grundlagen des dritten internationalen Vergleichs. In M. Prenzel, C. Artelt, J. Baumert, W. Blum, M. Hammann, E. Klieme & R. Pekrun (Hrsg.), *Pisa 2006. Die Ergebnisse der dritten internationalen Vergleichsstudie* (S. 367-380). Münster: Waxmann.
- Chavous, T. M., Bernat, D. H., Schmeelk-Cone, K. H., Caldwell, C. H., Kohn-Wood, L. P., & Zimmerman, M. A. (2003). Racial identity and academic attainment among African American adolescents. *Child Development*, 74(4), 1076-1090.
- Erikson, E. (1968). *Identity: Youth and crisis*. New York: Norton.
- Esser, H. (2006). *Sprache und Integration: Die sozialen Bedingungen und Folgen des Spracherwerbs von Migranten*. Frankfurt a.M.: Campus.
- Esser, H. (2009). Pluralisierung oder Assimilation. Effekte der multiplen Inklusion auf die Integration von Migranten. *Zeitschrift für Soziologie*, 38(5), 358-378.
- Fend, H. (2005). *Entwicklungspsychologie des Jugendalters*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Fulgini, A., Witkow, M., & Garcia, C. (2005). Ethnic identity and the academic adjustment of adolescents from Mexican, Chinese, and European backgrounds. *Developmental Psychology*, 41(5), 799-811.
- Graves, T. (1967). Psychological acculturation in a tri-ethnic community. *South-Western Journal of Anthropology*, 23(4), 337-350.
- Hannover, B., Morf, C., Neuhaus, J., Rau, M., Wolfgramm, C., & Zander-Music, L. (2013). How immigrant adolescents' self-views in school and family context relate to academic success in Germany. *Journal of Applied Social Psychology*, 43(1), 175-189.
- Horenczyk, G. (2008). Cultural identities, perceived discrimination, and adaption: Immigrant adolescents in Israel. In A. Stavans & I. Kupferberg (Hrsg.), *Studies in language and language education. Essays in honor of Elite Olshtain* (S. 395-410). Jerusalem: Magnes Press.
- Horenczyk, G. (2010). Language and identity in the school adjustment of immigrant students in Israel. *Zeitschrift für Pädagogik*, 55. Beiheft, 44-58.
- Klieme, E., Neubrand, M., & Lüdtke, O. (2001). Mathematische Grundbildung: Testkonzeption und Ergebnisse. In J. Baumert, E. Klieme, M. Neubrand, M. Prenzel, U. Schiefele, W. Schneider, P. Stanat, K. Tillmann & M. Weiß (Hrsg.), *PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich* (S. 139-190). Opladen: Leske + Budrich.
- Lewin, K. (1948). *Resolving social conflicts*. New York: Harper.
- Maehler, D. B. (2012). *Akkulturation und Identifikation bei eingebürgerten Migranten in Deutschland* (Reihe Internationale Hochschulschriften). Münster: Waxmann.

- Marcia, J., Waterman, A., Matteson, D., Archer, S., & Orlosky, J. (1993). *Ego identity: A handbook of psychological research*. New York: Springer.
- Möller, J., & Trautwein, U. (2009). Selbstkonzept. In E. Wild & J. Möller (Hrsg.), *Pädagogische Psychologie* (S. 179-204). Heidelberg: Springer Medizin.
- Müller, A. G., & Stanat, P. (2006). Schulischer Erfolg von Schülerinnen und Schülern mit Migrationshintergrund: Analysen zur Situation von Jugendlichen aus der ehemaligen Sowjetunion und der Türkei. In J. Baumert, P. Stanat & R. Watermann (Hrsg.), *Herkunftsbedingte Disparitäten im Bildungswesen: Differenzielle Bildungsprozesse und Probleme der Verteilungsgerechtigkeit* (S. 221-256). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Naumann, J., Artelt, C., Schneider, W., & Stanat, P. (2010). Lesekompetenz von PISA 2000 bis PISA 2009. In E. Klieme, C. Artelt, J. Hartig, N. Jude, O. Köller, M. Prenzel, W. Schneider & P. Stanat (Hrsg.), *PISA 2009. Bilanz nach einem Jahrzehnt* (S. 23-71). München: Waxmann.
- OECD (2009). *PISA 2006. Technical Report*. Paris: OECD.
- OECD (2010). *PISA 2009 Results: Overcoming social background. Equity in learning opportunities and outcomes (Volume II)*. Paris: OECD.
- OECD (2011). *PISA 2009 Ergebnisse: Potentiale nutzen und Chancengerechtigkeit sichern – Sozialer Hintergrund und Schülerleistungen (Band II)*. Paris: OECD.
- Oysermann, D., Harrison, K., & Bybee, D. (2001). Can racial identity be promotive of academic efficacy? *International Journal of Behavioral Development*, 25(4), 379-385.
- Oysermann, D., Kimmelmeier, M., Fryberg, S., Brosh, H., & Hart-Johnson, T. (2003). Racial-ethnic self-schemas. *Social Psychology Quarterly*, 66(4), 333-347.
- Phinney, J. S. (1990). Ethnic identity in adolescents and adults: Review of research. *Psychological Bulletin*, 108(3), 499-514.
- Phinney, J. S. (1992). The Multigroup Ethnic Identity Measure: A new scale for use with diverse groups. *Journal of Adolescent Research*, 7(2), 156-176.
- Phinney, J. S., Horenczyk, G., Liebkind, K., & Vedder, P. (2001). Ethnic identity, immigration, and well-being: An interactional perspective. *Journal of Social Issues*, 57(3), 493-510.
- Phinney, J. S., Berry, J. W., Vedder, P., & Liebkind, K. (2006). The acculturation experience: Attitudes, identities, and behaviors of immigrant youth. In J. W. Berry, J. S. Phinney, D. L. Sam & P. Vedder (Hrsg.), *Immigrant youth in cultural transition: Acculturation, identity and adaptation across national contexts* (S. 1-14). Mahwah: Lawrence Erlbaum Associates Publishers.
- Phinney, J. S., & Ong, A. D. (2007). Conceptualization and measurement of ethnic identity: Current status and future directions. *Journal of Counseling Psychology*, 54(3), 271-281.
- Portes, A., & Zhou, M. (1993). The new second generation: Segmented assimilation and its variants. *Annals of the American Academy of Political and Social Science*, 530(1), 74-96.
- Portes, A., & Rumbaut, R. G. (2001). Not everyone is chosen: Segmented assimilation and its determinants. In A. Portes & R. G. Rumbaut (Hrsg.), *Legacies: The story of the immigrant second generation* (S. 44-69). Berkeley: University of California Press.
- Prenzel, M., Carstensen, C. H., Frey, A., Drechsel, B., & Rönnebeck, S. (2007). PISA 2006. Eine Einführung in die Studie. In M. Prenzel, C. Artelt, J. Baumert, W. Blum, M. Hammann, E. Klieme & R. Pekrun (Hrsg.), *Pisa 2006. Die Ergebnisse der dritten internationalen Vergleichsstudie* (S. 31-60). Münster: Waxmann.
- Rudmin, F. W., & Ahmadzadeh, V. (2001). Psychometric critique of acculturation psychology: The case of Iranian migrants in Norway. *Scandinavian Journal of Psychology*, 42, 41-56.
- Rumbaut, R. G. (1994). The crucible within: Ethnic identity, self-esteem, and segmented assimilation among children of immigrants. *International Migration Review*, 28(4), 748-794.
- Schofield, J. W. (2006). *Migrationshintergrund, Minderheitenzugehörigkeit und Bildungserfolg. Forschungsergebnisse der pädagogischen, Entwicklungs- und Sozialpsychologie (Arbeitsstelle Interkulturelle Konflikte und gesellschaftliche Integration AKI)*. Berlin: Wissenschaftszentrum Berlin für Sozialforschung (WZB).

- Stanat, P., Rauch, D., & Segeritz, M. (2010). Schülerinnen und Schüler mit Migrationshintergrund. In E. Klieme, C. Artelt, J. Hartig, N. Jude, O. Köller, M. Prenzel, W. Schneider & P. Stanat (Hrsg.), *Pisa 2009. Bilanz nach einem Jahrzehnt* (S. 200-230). München: Waxmann.
- Tajfel, H., & Turner, J. C. (1979). An integrative theory of intergroup conflict. In W. G. Austin & S. Worchel (Hrsg.), *The social psychology of intergroup relations* (S. 33-47). Monterey: Brooks/Cole.
- Trickett, E., & Birman, D. (2005). Acculturation, school context, and school outcomes: Adaptation of refugee adolescents from Former Soviet Union. *Psychology in the Schools*, 42(1), 27-38.
- van der Linden, W., & Hambleton, R. K. (Hrsg.) (1997). *Handbook of modern item response theory*. New York: Springer.
- Vedder, P., Horenczyk, G., & Liebkind, K. (2006). Ethno-culturally diverse education settings: Problems, challenges and solutions. *Educational Research Review*, 1(2), 157-168.
- Vedder, P., van de Vijver, F., & Liebkind, K. (2006). Predicting immigrant youths' adaptation across countries and ethnocultural groups. In J. W. Berry, J. S. Phinney, D. L. Sam & P. Vedder (Hrsg.), *Immigrant youth in cultural transition: Acculturation, identity and adaptation across national contexts* (S. 143-166). London: Lawrence Erlbaum Associate Publishers.
- Ward, C. (1996). *Acculturation*. In D. Landis & R. S. Bhagat (Hrsg.), *Handbook of Intercultural Training* (S. 124-147). Thousand Oaks: Sage.
- Ward, C., & Kennedy, A. (1993). Psychological and sociocultural adjustment during cross-cultural transitions: A comparison of secondary students overseas and at home. *International Journal of Psychology*, 28(4), 129-147.
- Wong, E., Eccles, J., & Sameroff, A. (2003). The influence of ethnic discrimination and ethnic identification on African American adolescents' school and socioemotional adjustment. *Journal of Personality*, 71(6), 1197-1232.
- Zhou, M. (1997). Segmented assimilation: Issues, controversies, and recent research in the new second generation. *International Migration Review*, 31(4), 975-1008.

Anschrift der Autorinnen/des Autors

Dipl. Psych. Aileen Edele, Institut zur Qualitätsentwicklung im Bildungswesen (IQB),
Humboldt-Universität zu Berlin, Unter den Linden 6, 10099 Berlin, Deutschland
E-Mail: aileen.edele@iqb.hu-berlin.de

Prof. Dr. Petra Stanat, Institut zur Qualitätsentwicklung im Bildungswesen (IQB),
Humboldt-Universität zu Berlin, Unter den Linden 6, 10099 Berlin, Deutschland
E-Mail: petra.stanat@iqb.hu-berlin.de

B.A. Susanne Radmann, Institut zur Qualitätsentwicklung im Bildungswesen (IQB),
Humboldt-Universität zu Berlin, Unter den Linden 6, 10099 Berlin, Deutschland
E-Mail: susanne.radmann@iqb.hu-berlin.de

M.A. Michael Segeritz, Research Alliance for New York City Schools, New York University,
285 Mercer Street, New York, USA
E-Mail: michael.segeritz@nyu.edu

*Robin Segerer/Alexandra Marx/Petra Stanat/Wolfgang Schneider/
Thorsten Roick/Peter Marx*

Determinanten der Lesekompetenz bei Jugendlichen nicht deutscher Herkunftssprache

Zur Bedeutung der Spracherwerbsreihenfolge im Falle von Mehrsprachigkeit

1. Einleitung

1.1 Migrationshintergrund und Lesekompetenz

Obwohl sich die Leistungen von Schülerinnen und Schülern mit Migrationshintergrund im Rahmen des „Programme for International Student Assessment“ (PISA) in den letzten Jahren bedeutsam verbessert haben, scheinen in dieser Gruppe v.a. im Bereich der Lesekompetenz erhebliche Schwächen zu persistieren (Stanat, Rauch & Segeritz, 2010). Bei PISA 2009 erreichten mehr als 30% der Jugendlichen mit Migrationshintergrund nicht die PISA-Lesekompetenzstufe II (Naumann, Artelt, Schneider & Stanat, 2010). Entsprechend der kriterialen Interpretation der PISA-Kompetenzstufen war somit fast ein Drittel der Jugendlichen mit Migrationshintergrund nicht in der Lage, einfachste Schlussfolgerungen aus Textinformationen zu ziehen, die nicht explizit als relevant markiert worden waren. Bei den Jugendlichen ohne Migrationshintergrund betrug der entsprechende Anteil an schwachen Lesern gerade einmal 14% (Naumann et al., 2010). Neben den ungünstigeren soziokulturellen und materiellen Ausgangsbedingungen der Migrantenfamilien in Deutschland (Stanat et al., 2010) ist, Ergebnissen aus PISA und Studien über frühe Lesefertigkeiten zufolge, die in der Familie gesprochene deutsche Sprache entscheidend für die Ausprägung von Lesekompetenz im Deutschen verantwortlich (Stanat & Christensen, 2006; Walter, 2008; Tiedemann & Billmann-Mahecha, 2007; Stanat et al., 2010). Vor diesem Hintergrund erscheint es dringend geboten, den Faktor Mehrsprachigkeit, d.h. den ungesteuerten Erwerb und die fakultative, alltägliche Nutzung mehrerer Sprachen, differenzierter zu betrachten, um auf diese Weise einen vertieften Einblick in das Bedingungsgefüge der schwachen Lesekompetenz von Jugendlichen mit Migrationshintergrund zu erhalten.

1.2 Determinanten der Lesekompetenz

Determinanten des Leseverstehens sind kognitive Fähigkeiten wie Inhaltswissen, Wissen über Textmerkmale und Lesestrategien sowie fluide Intelligenz und Arbeitsgedächtnis. Darüber hinaus gelten grundlegende sprachliche Kompetenzen wie Wortschatz und Grammatik als Voraussetzung von Lesekompetenz, ebenso phonologische Bewusstheit

und andere metasprachliche Fähigkeiten, die basalen Dekodierleistungen zugrunde liegen (vgl. Artelt et al., 2007).

In vielen Studien werden die sich im Verlauf der Grundschulzeit entwickelnden Leseschwierigkeiten von Kindern, die die Verkehrssprache als Zweitsprache erlernt haben, auf Rückstände im Wortschatz zurückgeführt (z.B. Verhoeven, 2000). Der Wortschatz der Verkehrssprache fällt bei multilingualen Probanden gemeinhin geringer aus als bei monolingualen Personen (Bialystok, 2010). Kinder, die nicht in der Verkehrssprache aufgewachsen sind, weisen in der Grundschule ebenfalls anhaltende Schwächen in grammatikalischen Bereichen auf (vgl. z.B. Chiappe, Siegel & Wade-Woolley, 2002). Was die Fähigkeitsausprägungen im Bereich der phonologischen Bewusstheit betrifft, so zeichnet sich kein einheitliches Bild ab. Weber, Marx und Schneider (2007) zeigten auf, dass das Erlernen der Verkehrssprache als Zweitsprache mit geringeren Werten im Bereich der phonologischen Bewusstheit im Deutschen einhergeht. Limbird und Stanat (2006) hingegen konnten bei deutsch-türkischen Kindern und sprachneutralem Stimulusmaterial keine signifikanten Unterschiede im Bereich der phonologischen Bewusstheit nachweisen. Mitunter zeigten sich Schwächen der Zweitsprachler nur in Tests der phonologischen Bewusstheit mit real vorkommenden Wörtern als Stimulusmaterial, d.h. in Aufgaben, die mit Anforderungen an den Wortschatz konfundiert waren (z.B. Chiappe et al., 2002). Darüber hinaus gibt es sogar substantielle Belege dafür, dass Multilingualität mit einem erhöhten Maß an phonologischer Bewusstheit einhergehen kann (vgl. Bialystok, Majumder & Martin, 2003).

1.3 Bedeutsamkeit der Spracherwerbsbiografie

In den meisten Studien zu basalen Determinanten der Lesekompetenz wird nicht deutlich, ob die Nachteile multilingualer Schülerinnen und Schüler auf ihre Mehrsprachigkeit, d.h. auf das kompetitive Verhältnis der Sprachen untereinander, zurückzuführen sind oder darauf, dass die Verkehrssprache als Zweitsprache, d.h. später als die Erstsprache erworben wurde. Es ist hinlänglich bekannt, dass die Fähigkeitsausprägung in verschiedenen sprachlichen Bereichen umso geringer ausfällt, je später mit dem Erwerb einer zweiten Sprache begonnen wurde (z.B. Bialystok & Hakuta, 1999). In differenzierten Analysen der ersten beiden PISA-Zyklen zeigte sich ein signifikanter Effekt des Immigrationszeitpunktes der Familie auf die Lesekompetenz. Waren die Schülerinnen und Schüler mit Migrationshintergrund erst nach oder während der Grundschulzeit eingewandert, so ging dies mit gravierenden zusätzlichen, d.h. über den Effekt des Migrationshintergrundes hinausgehenden Lesekompetenzrückständen einher (Walter, 2008).

Im Rahmen der DESI-Studie 2006 (Hesse, Göbel & Hartig, 2008) zeigten sich beträchtliche Unterschiede zwischen simultan mehrsprachig aufgewachsenen Jugendlichen und Jugendlichen mit Deutsch als Zweitsprache (DaZ). Probanden mit Deutsch als Zweitsprache schnitten in den Bereichen Lesekompetenz, Wortschatz, grammatikalische Sprachbewusstheit, Rechtschreibung und schriftsprachliche Pragmatik schlechter ab als monolinguale Schülerinnen und Schüler und auch schlechter als simultan mehr-

sprachig aufgewachsene Probanden. Die Effekte erscheinen umso eindrucksvoller, da die Autoren annehmen, dass ein bedeutender Anteil der DaZ-Gruppe bereits im frühen Kindesalter mit dem Erwerb der deutschen Sprache begonnen hatte (Hesse et al., 2008). Simultan mehrsprachig aufgewachsene Schülerinnen und Schüler wiesen nur im Wortschatz bedeutsame Schwächen gegenüber rein monolingual aufgewachsenen Probanden auf.

1.4 Effekte von Erwerbsdauer oder Erwerbszeitpunkt?

Der Zweitspracherwerb geht im Vergleich zum Erstspracherwerb nicht nur mit einer kürzeren Dauer des Sprachkontaktes einher, sondern ab bestimmten Erwerbszeitpunkten auch mit verringerter neuronaler Plastizität bezüglich der verschiedenen sprachlichen Bereiche (vgl. Bialystok & Hakuta, 1999).

Von der Dauer des Sprachkontaktes scheinen vornehmlich semantische Determinanten der Lesekompetenz abzuhängen, wie etwa der Wortschatz (Cummins, 1981). Für die formalen sprachlichen Bereiche existieren hingegen Befunde, die dafür sprechen, dass ein sehr früher Spracherwerbszeitpunkt notwendig ist, um eine muttersprachliche Kompetenzausprägung zu erreichen (Weber-Fox & Neville, 1996). Klare Belege für eine solche frühe kritische Phase gibt es im Bereich der rezeptiven Phonologie (Kuhl, 2004). Eine exakte Phonemwahrnehmung ist nicht nur eine notwendige Voraussetzung von phonologischer Bewusstheit, sondern ermöglicht darüber hinaus das Erkennen von Wortgrenzen innerhalb des Lautstroms gesprochener Sprache, somit Hörverstehen sowie den impliziten Erwerb von Wortschatz und den Aufbau grammatikalischer Strukturen (Saffran, Aslin & Newport, 1996; Bates & Goodman, 1998). Als ähnlich bedeutsam hat sich die Sensitivität für phonotaktische oder prosodische Segmentierungshinweise erwiesen, die von Zweitsprachlern häufig nicht genutzt werden können (Penner, 2005; Cutler, 2001). Auch im Bereich der grammatikalischen Fähigkeiten ließ sich aufzeigen, dass bereits leichte Verzögerungen des Grammatikerwerbs bedeutsame spätere Kompetenznachteile zur Folge haben können (Weber-Fox & Neville, 1996; Wartenburger et al., 2003).

Es ist verwunderlich, dass nur wenig darüber bekannt ist, ob sich der konkrete Erwerbszeitpunkt des Sprachkontaktes vermittelt über Schwierigkeiten in Bereichen der formalen Sprache auf die Lesekompetenzentwicklung auswirkt. Lediglich in einer Studie mit spanisch-englischsprachigen Kindern fanden sich Hinweise auf die Bedeutsamkeit eines frühzeitigen Spracherwerbsbeginns für frühe Lesefertigkeiten in der Grundschule (Kovelman, Baker & Petitto, 2008). Generell wird angenommen, dass Schwierigkeiten in formalen Bereichen der Sprache durch ein erhöhtes Maß an kognitiver Kontrolle kompensiert werden müssen (Wartenburger et al., 2003). Daher erscheint es möglich, dass sich Schwächen in formalen Bereichen der Zweitsprache v.a. bei Probanden mit geringer ausgeprägten kognitiven Kontrollmöglichkeiten bedeutsam auf das Leseverstehen auswirken.

2. Forschungsfragen

Die vorliegende Untersuchung im Rahmen der DELKO-Studie (Determinanten der Lesekompetenz)¹ soll Fähigkeitsunterschiede in basalen semantischen und formalen sprachlichen Kompetenzbereichen zwischen leseschwachen mehrsprachigen Jugendlichen mit Deutsch als Zweitsprache (DaZ), leseschwachen mehrsprachigen Jugendlichen mit Deutsch als Erstsprache (DaE) sowie deutsch-monolingualen leseschwachen Jugendlichen nachweisen. Darüber hinaus soll untersucht werden, ob die unterschiedlichen Fähigkeitsausprägungen in diesen Bereichen mit Unterschieden im Bereich der Lesekompetenz einhergehen. Die Differenzierung in mehrsprachige Jugendliche mit DaE und DaZ lehnt sich eng an die Konzeption der DESI-Studie an (Hesse et al., 2008). Im Gegensatz zur DESI-Studie bezieht sich die vorliegende Studie jedoch vornehmlich auf Leserinnen und Leser, die im unteren Leistungsspektrum zu verorten sind und von denen anzunehmen ist, dass ihnen in deutlich geringerem Umfang kognitive Kompensationsmöglichkeiten zur Verfügung stehen. Standen bei DESI relativ elaborierte linguistische und metalinguistische Fähigkeiten im Vordergrund, so sollen in dieser Untersuchung das Spektrum möglicher Determinanten erweitert und neben Wortschatz sowie grammatikalischen Fähigkeiten auch basalere Kompetenzen wie phonologische Bewusstheit und prosodische Sensitivität untersucht werden. Kompetenzunterschiede zwischen den mehrsprachigen Jugendlichen mit DaE und DaZ sollten sich in allen sprachlichen Bereichen und nicht zuletzt auch in der Lesekompetenz aufzeigen lassen. Fähigkeiten im Bereich der formalen Sprache, d.h. phonologische Bewusstheit, prosodische Sensitivität und Grammatikkompetenz, scheinen von einem möglichst frühen Beginn des Spracherwerbs abzuhängen, daher müssten hier gravierende Schwächen bei mehrsprachigen Probanden mit DaZ gegenüber mehrsprachigen Probanden mit DaE und auch gegenüber monolingualen Leserinnen und Lesern festzustellen sein.

Im semantischen Bereich des Wortschatzes ist zu erwarten, dass beide mehrsprachigen Gruppen schlechter abschneiden als die monolinguale Vergleichsgruppe, wobei die mehrsprachige Gruppe mit DaZ aufgrund des geringeren Sprachkontaktes ebenfalls geringere Werte als die mehrsprachige Gruppe mit DaE aufweisen sollte.

3. Methodik

3.1 Stichprobe

Für die DELKO-Studie wurde ergänzend zur PISA-Erhebung des Jahres 2009 eine zusätzliche Stichprobe von Schülerinnen und Schülern rekrutiert, bei denen von einem ho-

1 Das Hauptziel der DELKO-Studie ist es, differenzierte Kompetenzprofile von leseschwachen Schülerinnen und Schülern anhand verschiedener sprachlicher und nicht sprachlicher kognitiver Fähigkeitsindikatoren und unter Berücksichtigung des sprachlichen Hintergrunds zu erstellen.

hen Anteil schwacher Leserinnen und Leser ausgegangen werden konnte. Die Stichprobenziehung erfolgte in mehreren deutschen Großstädten, an 20 Haupt-, Real- und Gesamtschulen, die einen Anteil an Jugendlichen mit Migrationshintergrund von mehr als 40% aufwiesen. Die Erfassung der Spracherwerbsreihenfolge basierte auf einer Selbstangabe der Schülerinnen und Schüler in einem Fragebogen (*Welche Sprache hast du in deiner Familie zuerst gelernt [Erstsprache/Muttersprache]?*). Die Schülerinnen und Schüler konnten dabei auch angeben, dass sie mehrere Sprachen gleichzeitig als Muttersprache erworben hatten. Zwei weitere Fragen zur exakten zeitlichen Beschreibung der Spracherwerbsbiografie erwiesen sich als wenig verlässlich und wurden daher nur dazu verwendet, fehlende Angaben aufzufüllen. Schülerinnen und Schüler mit Migrationshintergrund, von denen keinerlei Hinweise auf die Spracherwerbsbiografie vorlagen, wurden aus der Studie ausgeschlossen.

Es ergab sich eine Gesamtstichprobe von 410 Neuntklässlern, bestehend aus 107 deutsch-monolingualen Jugendlichen (im Weiteren als monolingual bezeichnet), bei denen kein Anhaltspunkt vorlag, dass innerhalb ihrer Herkunftsfamilie eine weitere Sprache erworben wurde, 105 mehrsprachigen Jugendlichen mit Deutsch als Erstsprache (DaE), die Deutsch entweder ausschließlich oder zusammen mit einer anderen Sprache als Erstsprache erworben hatten, und 198 Jugendlichen mit Deutsch als Zweitsprache (DaZ).

3.2 Operationalisierung

Die Messung der Fähigkeitsbereiche phonologische Bewusstheit und prosodische Sensitivität sowie Grammatik erfolgte computerbasiert mithilfe von Laptops, die in einem mobilen Netzwerk eingebunden waren. Die Aufgabenzuteilung konnte auf diese Weise synchronisiert und die Daten konnten zentral gespeichert werden. Alle weiteren Aufgaben wurden als papiergestützte Testverfahren administriert.

Lesekompetenz

Die Messung der Lesekompetenz der Schülerinnen und Schüler erfolgte über die international standardisierte Leistungsmessung der PISA-Studie und beinhaltete die drei Subskalen „Informationen suchen und extrahieren“, „textbezogenes Kombinieren und Interpretieren“ sowie „Reflektieren und Bewerten“ (vgl. Naumann et al., 2010). Die PISA-Leseitems bestehen aus einer Mischung aus Mehrfachwahlaufgaben und Fragen mit freiem Antwortformat. Alle Items sind anhand eines übergeordneten Textstimulus (z.B. Passagen aus einem Theaterstück, Diagramme, Zeitschriftenartikel o.Ä.) zu beantworten. Da die PISA-Items raschhomogene Eigenschaften aufweisen, konnte die Erfassung der Fähigkeiten der Schülerinnen und Schüler mithilfe von 13 verschiedenen Testversionen mit variierenden Aufgaben erfolgen (Multi-Matrix-Design). Im Durchschnitt bearbeiteten die Schülerinnen und Schüler 31.08 Aufgaben ($SD = 13.00$). Die verwendeten Booklets gehörten zu einer besonderen Auswahl einfacherer Testhefte (sog. „easy

sets“), die speziell dafür entwickelt wurden, um im unteren Fähigkeitsbereich eine bessere Differenzierung und damit eine zuverlässigere Erfassung zu ermöglichen (Nauermann et al., 2010). Die biasfreie Schätzung der Lesekompetenzwerte erfolgte unter Zuhilfenahme eines Hintergrundmodells, bestehend aus allen Einzelvariablen und Skalen, die auch im internationalen Hintergrundmodell verwendet werden, den nationalspezifisch gebildeten Skalen aus Schüler- und Elternfragebogen sowie ausgewählten Skalen aus dem Lehrerfragebogen (vgl. OECD, 2009). Zusätzlich wurden die in DELKO erfassten kognitiven und sprachlichen Variablen berücksichtigt. Die Schätzung der Lesekompetenz erfolgte in Form von jeweils fünf „plausible values“ für jede der drei Subdimensionen des Lesens (vgl. Wu, 2005; von Davier, Gonzalez & Mislevy, 2009).

Phonologische Bewusstheit

Die 12 Items zur Erfassung der phonologischen Bewusstheit im engeren Sinne erfordern die Identifikation eines Wortes innerhalb einer Gruppe von jeweils vier akustisch dargebotenen einsilbigen Pseudowörtern, die alle aus einer Konsonant-Vokal-Konsonant-Kombination (KVK) gebildet wurden. Bei Items zur *Mittellautidentifikation* unterscheidet sich der mittlere Vokal des zu identifizierenden Wortes von dem der anderen Wörter. Bei der *Endlautidentifikation* weicht der Endkonsonant eines der Pseudowörter von dem der übrigen Wörter ab (*Beispielitem*: $\square pat \square kut \square pit \square pal$). Die Probanden mussten angeben, an welcher Stelle sich das gesuchte Wort innerhalb der Wortreihe befindet. Das Verfahren wurde neu entwickelt und vorab an einer Stichprobe von annähernd 400 Jugendlichen an Hauptschulen in Berlin und Bayern pilotiert. Die interne Konsistenz des Tests kann mit $\alpha = .70$ als zufriedenstellend bewertet werden.

Prosodische Sensitivität

Der Test zur prosodischen Sensitivität beinhaltet 15 Items, die ebenfalls aus vier akustisch dargebotenen einsilbigen KVK-Pseudowörtern bestehen. In dieser Aufgabe wird die Quantität des Mittellauts verringert oder erhöht, d.h. die Vokallänge bei einem der Pseudowörter gegenüber denen der anderen drei Wörter verlängert oder verkürzt. Dieses Wort sollte jeweils identifiziert werden (*Beispielitem*: $\square maar \square raas \square dack \square laa$). Die Items stammen aus einer vorläufigen Version des Basiskompetenztests 5-9 (BAKO; Stock & Schneider, in Vorb.). Die interne Konsistenz der Skala ist mit $\alpha = .78$ zufriedenstellend bis gut.

Grammatikkompetenz

Der Grammatiktest besteht aus 29 Mehrfachwahlaufgaben zum korrekten Umgang mit Flektions-, Verbindungs- und Wortbildungsmorphemen unter Berücksichtigung verschiedener syntaktischer Cues und Rahmenstrukturen. Der erste Teil des Tests fokussiert die Nutzung basaler unregelmäßiger Konjugations- und Deklinationsformen sowie das Beherrschen wenig salienter Strukturen wie Funktionswörter (z.B. Präpositionen)

und deren korrekte syntaktische Einbettung (*Beispielitem: Er betonte immer wieder, dass er schuld sei _____ all der Ungerechtigkeit. Antwortmöglichkeiten: ☐ an ☐ von ☐ um ☐ vor*). Der zweite Teil des Tests, ca. ein Drittel der Items, erfolgt unter Verwendung von Pseudowörtern, was die Verwendung allzu unregelmäßiger grammatikalischer Strukturen unmöglich macht (*Beispielitem: Ich bin in meinem Leben nie _____ worden. Antwortmöglichkeiten: ☐ plosen ☐ ploster ☐ geplostet ☐ verploster*). Dieses Vorgehen soll eine Differenzierung zwischen einer holistisch automatisierten Verarbeitung und einer expliziten metalinguistisch kontrollierten Verarbeitung in den Aufgaben mit Pseudowörtern ermöglichen (vgl. auch Kuo & Anderson, 2008). Das Testverfahren wurde schriftlich und ohne Zeitbegrenzung vorgegeben, um die Grammatikkompetenzen unabhängig von der phonologischen Informationsverarbeitungsgeschwindigkeit erfassen zu können. Das Verfahren wurde neu entwickelt und vorab an einer Stichprobe von annähernd 400 Jugendlichen an Hauptschulen in Berlin und Bayern pilotiert. Die interne Konsistenz der Gesamtskala beträgt $\alpha = .85$. Das Verfahren weist somit einen guten Homogenitätswert auf.

Wortschatz

Die Erfassung des Wortschatzes erfolgte durch den CFT-20-R/WS, Version A (Weiß, 2008), durch den der über den Grundwortschatz der deutschen Sprache hinausgehende rezeptive Wortschatz erfasst werden kann. Der Test setzt sich aus 30 Mehrfachwahlaufgaben zusammen, in denen jeweils ein Begriff vorgegeben wird und ein synonyme Begriff unter fünf vorgegebenen Wörtern gefunden werden muss. Die Skala weist mit $\alpha = .95$ eine sehr gute interne Konsistenz auf.

Nonverbale Intelligenz

Die nonverbale Intelligenz der Schüler wurde über die Subskala N2 „Figurenanalogien“ des KFT 4-12+R (Kognitiver Fähigkeitstest für 4. bis 12. Klassen, Revision; Heller & Perleth, 2000), Testversion B, erfasst. Zwei fehlerhafte Items wurden aus dem Testheft entfernt (vgl. hierzu Segerer, Marx & Marx, im Druck). Die Skala weist mit $\alpha = .92$ eine gute interne Konsistenz auf.

Sozioökonomischer-soziokultureller Hintergrund

Der sozioökonomische-soziokulturelle Hintergrund der Jugendlichen wurde mittels des ESCS-Indexes (*Index of educational, social, and cultural status*) erfasst (vgl. Ehmke & Siegle, 2005). Der Index integriert drei Subkomponenten: den höchsten Bildungsabschluss und den höchsten Berufsstatus in der Familie sowie die ökonomischen und kulturellen Ressourcen der Familie. Diese letzte Komponente berücksichtigt Angaben über den Besitz von verschiedenen bildungsrelevanten Gütern wie Computer oder Bücher, die Anzahl an Schlafräumen sowie Bildungsgelegenheiten wie den Besuch von Museen und Bibliotheken. Auf diese Weise destilliert der Index das verfügbare ökonomische

mische und kulturelle Kapital auf ein nutzbares quantitatives Maß. Der Index wurde anhand des Mittelwerts der OECD-Population und deren Standardabweichung z-standardisiert (Ramm et al., 2006).

4. Statistische Analysen

4.1 Modellierung latenter Variablen

Zunächst wurden fehlende Werte in den einzelnen Skalen unter Einbezug aller relevanten Variablen mithilfe des Programms SPSS 18.0 multipel imputiert und fünf vollständige Datensets generiert (vgl. Lüdtke, Robitzsch, Trautwein & Köller, 2007). Die abhängige Variable Lesekompetenz sowie die möglichen erklärenden Variablen phonologische und prosodische Sensitivität, Grammatikkompetenz und Wortschatz wurden unter Verwendung des Programms Mplus 6.1 (Muthén & Muthén, 1998-2010) latent modelliert. Dies erfolgte einerseits, um die Messfehler der Instrumente zu bereinigen, andererseits, um die Messinvarianz der Tests über die verschiedenen Gruppen hinweg gewährleisten zu können. Da für die Fähigkeitsbereiche phonologische Bewusstheit und prosodische Sensitivität sowie Wortschatz jeweils ein homogener Test als manifeste Indikator zur Verfügung stand, erfolgte die Modellierung der latenten Variablen analog zu einem „Odd-even-Parceling“ (vgl. Kim & Hagtvet, 2003) durch alternierende Zuteilung der Testitems zu drei Itempaketen. Zur latenten Modellierung der Grammatikkompetenz fand dieses Parceling-Verfahren sowohl für den Testteil mit als auch ohne Pseudowörter Verwendung, sodass insgesamt sechs Itempakete gebildet wurden. Der metalinguistische Mehraufwand, den der Umgang mit Pseudowörtern erfordert, wurde mittels eines indikatorspezifischen metalinguistischen Methodenfaktors modelliert (Gustafsson & Balke, 1993). Die latente Modellierung des Lesekompetenzwerts erfolgte über die drei PISA-Subskalen „Informationen suchen und extrahieren“, „textbezogenes Kombinieren und Interpretieren“ sowie „Reflektieren und Bewerten“ (vgl. Naumann et al., 2010).

Alle verwendeten Verfahren wurden innerhalb eines Gesamtmodells mit Lesekompetenz als Kriterium und den weiteren latenten Variablen als Prädiktoren unter regressionsanalytischer Kontrolle der leserelevanten manifesten Kovariaten Geschlecht, non-verbale Intelligenz sowie sozioökonomischer-soziokultureller Hintergrund zunächst auf konfigurale, metrische und skalare Messinvarianz geprüft (Vandenberg & Lance, 2000). Konfigurale Invarianz bezeichnet dabei den Nachweis, dass über alle Gruppen hinweg dieselbe Faktorenstruktur auf Basis der manifesten Indikatoren modelliert wird. Der Nachweis metrischer Invarianz belegt darüber hinausgehend, dass die Faktorladungen der manifesten Indikatoren über alle Gruppen hinweg gleich sind, und skalare Invarianz liegt dann vor, wenn die Intercepts der jeweiligen Indikatoren über alle Gruppen hinweg ebenfalls gleich groß geschätzt werden können. Die Annahme dieser Formen der Invarianz kann dann erfolgen, wenn das jeweils restriktivere Modell die vorliegende Kovarianzmatrix nicht bedeutsam schlechter abbildet als das jeweils weniger restriktive Modell (vgl. Vandenberg & Lance, 2000).

4.2 Analysestrategie

Zunächst werden unter regressionsanalytischer Kontrolle der manifesten Kovariaten Geschlecht, Intelligenz sowie sozioökonomischer-soziokultureller Hintergrund die latenten Mittelwertunterschiede (d.h. Unterschiede im Intercept der Regressionsgleichungen) zwischen monolingualen Jugendlichen, den mehrsprachigen Jugendlichen mit DaE und den mehrsprachigen Jugendlichen mit DaZ mittels Mehrgruppenstrukturgleichungsmodellen untersucht (vgl. Hancock, 2001). Zur Skalierung der latenten Variablen wird hierbei jeweils ein Gruppenmittelwert (Intercept) auf Null festgelegt. Die Mittelwerte der übrigen zwei Gruppen werden in Referenz zu diesem Nullpunkt betrachtet (Hancock & Mueller, 2006). Der Vergleich jeweils zweier Gruppen erfolgt mittels des Wald-Chi-Quadrat-Tests (vgl. Buse, 1982), der in Mplus 6.1 auch für Daten mit multiplen Imputationen verfügbar ist (Asparouhov & Muthén, 2010). Die Frage, welche der Erklärungsvariablen bedeutsame Gruppenunterschiede im Lesen erklären können, soll durch zusätzliche regressionsanalytische Kontrolle der latent modellierten Variablen phonologische Bewusstheit, prosodische Sensitivität, Wortschatz und Grammatikkompetenz beantwortet werden. Der Vergleich der Gruppenmittelwerte erfolgt dann analog zu den Vergleichen ohne Erklärungsvariablen mittels des Wald-Chi-Quadrat-Tests. Da phonologische Bewusstheit und prosodische Sensitivität als Voraussetzungen von Wortschatz und Grammatikkompetenz gelten (vgl. Kuhl, 2004) und der Wortschatz seinerseits die zentrale Determinante von Grammatikkompetenz darstellt (vgl. Bates & Goodman, 1998), kann die Bedeutung der hierarchieniedrigeren Fähigkeiten für die Aufklärung von Lesekompetenzunterschieden zwischen den Gruppen in einem Gesamtmodell nicht angemessen beurteilt werden. Es müssen daher zunächst Grammatikkompetenz als hierarchiehöchste Fähigkeit und anschließend Wortschatz aus dem Gesamtmodell ausgeschlossen werden. Der Vergleich der gruppenspezifischen Intercepts in den schrittweise reduzierten Regressionsmodellen erlaubt es, die Bedeutsamkeit des Wortschatzes und der basalen Bereiche phonologische Bewusstheit und prosodische Sensitivität für die Erklärung von Gruppenunterschieden im Lesen angemessen sichtbar zu machen.

5. Ergebnisse

5.1 Deskriptive Ergebnisse

Das Geschlechterverhältnis war in allen drei Stichproben ausgeglichen, im Alter unterschieden sich die Gruppen nicht signifikant, auch im Bereich der nonverbalen Intelligenz zeigten sich keine Gruppenunterschiede. Im Bereich des sozioökonomischen-soziokulturellen Hintergrunds schnitten Jugendliche mit DaZ signifikant schlechter ab als monolinguale Jugendliche (s. Tab. 1).

	Monolingual (n = 107)	Mehrsprachig/DaE (n = 105)	Mehrsprachig/DaZ (n = 198)
	M (SD)	M (SD)	M (SD)
Alter	15.97 _a (0.64)	15.96 _a (0.79)	16.09 _a (0.80)
ESCS	-0.43 _a (0.77)	-0.62 _{a,b} (1.01)	-0.72 _b (0.94)
KFT/N2	11.26 _a (6.03)	11.36 _a (5.36)	10.79 _a (5.65)
Geschlecht (weiblich)	42.06% _a	48.57% _a	46.46% _a

Anmerkungen: Mittelwerte und Prozentangaben mit ungleichen Fußindizes (a, b) unterscheiden sich mit $p < .05$ statistisch überzufällig.

Tab. 1: Stichprobenkennwerte der monolingualen Probanden, der mehrsprachigen Probanden mit DaE und der mehrsprachigen Probanden mit DaZ

		Monolingual (n = 107)	Mehrsprachig/DaE (n = 105)	Mehrsprachig/DaZ (n = 198)
	Itemanzahl	M (SD)	M (SD)	M (SD)
Lesen (global) ^a		427.71 _a (82.40)	407.16 _{a,b} (89.23)	396.98 _b (84.22)
Lesen (ISE) ^a		416.67 _a (100.34)	405.21 _a (110.50)	390.50 _a (105.57)
Lesen (tKI) ^a		435.22 _a (85.01)	414.97 _{a,b} (87.67)	401.15 _b (85.88)
Lesen (RB) ^a		417.65 _a (94.87)	393.14 _{a,b} (103.54)	385.38 _b (98.34)
Grammatik (gesamt) ^b	29	21.36 _a (4.53)	20.17 _{a,b} (5.29)	19.43 _b (5.20)
Grammatik/reale Wörter ^b	18	14.71 _a (2.58)	13.26 _b (3.36)	12.53 _b (3.56)
Grammatik/Pseudowörter ^b	11	6.65 _a (2.90)	6.91 _a (2.62)	6.90 _a (2.46)
Wortschatz ^b	30	23.92 _a (3.39)	22.66 _b (3.25)	21.74 _c (4.01)
Phonologie ^b	12	6.70 _a (2.73)	6.92 _a (2.46)	6.98 _a (2.53)
Prosodie ^b	15	7.78 _{a,b} (3.19)	8.34 _a (3.15)	7.51 _b (3.41)

Anmerkungen: ISE = Informationen suchen und extrahieren; tKI = textbezogenes Kombinieren und Interpretieren; RB = Reflektieren und Bewerten. a) Plausible Value; b) Summenwert. Mittelwerte und Prozentangaben mit ungleichen Fußindizes (a, b) unterscheiden sich mit $p < .05$ statistisch überzufällig.

Tab. 2: Deskriptive manifeste Ergebnisse der monolingualen Probanden, der mehrsprachigen Probanden mit DaE und der mehrsprachigen Probanden mit DaZ

Die durchschnittlichen globalen Lesefähigkeitsparameter der monolingualen Jugendlichen und der mehrsprachigen Jugendlichen mit DaE (s. Tab. 2) liegen am unteren Ende der Kompetenzstufe II, Jugendliche mit DaZ sind durchschnittlich noch der Kompetenzstufe Ia zuzuordnen (Naumann et al., 2010). Statistisch bedeutsam ist jedoch nur der Unterschied zwischen den plausible values der monolingualen Jugendlichen und denen der mehrsprachigen Jugendlichen mit DaZ. Die Mittelwerte und Standardabweichungen, bezogen auf die Anzahl an Items der verwendeten Skalen, belegen für alle verwendeten Verfahren hinreichende Differenzierungsfähigkeit innerhalb und zwischen den untersuchten Stichproben.

5.2 Messinvarianzprüfung

Die Überprüfung der Messinvarianz im Mehrgruppenregressionsmodell mit Lesekompetenz als Kriterium, mit allen latent modellierten Variablen als Prädiktoren und den leserelevanten manifesten Kontrollvariablen erbrachte keine Hinweise auf Einschränkungen im Hinblick auf konfigurale, metrische oder skalare Messinvarianz (s. Tab. 3).

Invarianz-Modell	χ^2/df	RMSEA	CFI	TLI	Vergleich mit	$\Delta \chi^2/df$
A: konfigural	386.11/494	<.01	>.99	>.99	–	–
B: metrisch	39.80/524	<.01	>.99	>.99	Modell A	13.69/26
C: skalar	418.58/548	<.01	>.99	>.99	Modell B	18.78/28

Anmerkungen: RMSEA = Root Mean Square Error of Approximation; CFI = Comparative Fit Index; TLI = Tucker-Lewis Index.

Tab. 3: Invarianzüberprüfung der latent modellierten Kriteriumsvariable Lesekompetenz und der latent modellierten Kovariaten phonologische Bewusstheit, prosodische Sensitivität, Wortschatz und Grammatikkompetenz

5.3 Ergebnisse der latenten Mittelwertvergleiche

Auf latenter Ebene fanden sich unter regressionsanalytischer Kontrolle der manifesten Kovariaten Geschlecht, Intelligenz sowie sozioökonomischem-soziokulturellem Hintergrund keine signifikanten Unterschiede zwischen den beiden Gruppen der mehrsprachigen Jugendlichen im Bereich der Lesekompetenz ($Wald-\chi^2(1) = 0.01$; $p = .92$). Beide mehrsprachigen Gruppen wiesen jedoch verglichen mit monolingual deutschen Jugendlichen (monolingual vs. DaE: $Wald-\chi^2(1) = 4.12$; $p = .04$; monolingual vs. DaZ: $Wald-\chi^2(1) = 4.80$; $p = .03$) signifikant geringere Lesekompetenzwerte auf (s. Tab. 4).

	Monolingual (<i>n</i> = 107)	Mehrsprachig/DaE (<i>n</i> = 105)	Mehrsprachig/DaZ (<i>n</i> = 198)
Intercepts	α (SE)	α (SE)	α (SE)
Lesekompetenz*	0.00 _a (–)	–0.64 _b (0.29)	–0.63 _b (0.30)
Phonologie	0.00 _a (–)	–0.01 _a (0.48)	0.01 _a (0.43)
Prosodie	0.00 _{a,b} (–)	0.38 _a (0.43)	–0.37 _b (0.33)
Wortschatz	0.00 _a (–)	–0.40 _a (0.48)	–0.88 _b (0.30)
Grammatik	0.00 _a (–)	–1.18 _b (0.29)	–1.04 _b (0.27)

Anmerkungen: Gerichtete Pfade der manifesten Kontrollvariablen Geschlecht, Intelligenz sowie sozioökonomischer-soziokultureller Hintergrund auf alle latenten Variablen. Der Mittelwert der monolingualen Probanden wurde zwecks Skalierung auf Null fixiert. Latente Mittelwerte mit ungleichen Fußindizes unterscheiden sich mit $p < .05$ statistisch überzufällig. Die Intercepts der Variablen phonologische Bewusstheit, prosodische Sensitivität, Wortschatz und Grammatikkompetenz sind dem Gesamtmodell (s. Tab. 5) entnommen. * Modell Lesekompetenz: $\chi^2/df = 43.19/34$, $p = .13$; RMSEA = .04; CFI = .99; TLI = .99.

Tab. 4: Latente Mittelwertunterschiede in den Fähigkeitsbereichen Lesekompetenz, phonologische Bewusstheit, prosodische Sensitivität, Wortschatz und Grammatikkompetenz

Im Bereich der phonologischen Bewusstheit unterschieden sich die mehrsprachigen Gruppen nicht voneinander ($Wald\text{-}\chi^2(1) = 0.00$; $p = .98$), und auch gegenüber der monolingualen Gruppe schnitten mehrsprachige Jugendliche mit DaE ($Wald\text{-}\chi^2(1) = 0.00$; $p = .99$) und DaZ ($Wald\text{-}\chi^2(1) = 0.00$; $p = .99$) vergleichbar ab.

Es fanden sich signifikante Differenzen zwischen den beiden Gruppen mehrsprachiger Jugendlicher unter Kontrolle von Geschlecht, Intelligenz, sozioökonomischem-soziokulturellem Hintergrund im Bereich der prosodischen Sensitivität ($Wald\text{-}\chi^2(1) = 4.64$; $p = .03$). Mehrsprachige Jugendliche mit DaE schnitten signifikant besser ab als mehrsprachige Jugendliche mit DaZ. Weder die mehrsprachige Gruppe mit DaE ($Wald\text{-}\chi^2(1) = 0.81$; $p = .37$) noch die mehrsprachige Gruppe mit DaZ ($Wald\text{-}\chi^2(1) = 1.15$; $p = .28$) wiesen von der monolingualen Gruppe unterscheidbare Prosodiewerte auf.

Im Hinblick auf den rezeptiven Wortschatz wies die mehrsprachige Gruppe mit DaZ signifikant geringere Werte auf als mehrsprachige Jugendliche mit DaE ($Wald\text{-}\chi^2(1) = 4.12$; $p = .04$). Im Vergleich zur monolingualen Gruppe wies nur die Gruppe der mehrsprachigen Jugendlichen mit DaZ einen signifikant geringeren Wortschatz auf ($Wald\text{-}\chi^2(1) = 8.41$; $p < .01$), die Gruppe der mehrsprachigen Jugendlichen mit DaE unterschied sich nicht bedeutsam von den monolingualen Jugendlichen ($Wald\text{-}\chi^2(1) = 0.68$; $p = .41$).

Hinsichtlich der Grammatikkompetenz unterschieden sich beide mehrsprachigen Gruppen nicht signifikant voneinander ($Wald\text{-}\chi^2(1) = 0.28$; $p = .60$). Gegenüber der mo-

nolingual deutschen Gruppe schnitten jedoch sowohl die mehrsprachigen Jugendlichen mit DaE ($Wald\text{-}\chi^2(1) = 14.70; p < .01$) als auch die mehrsprachigen Jugendlichen mit DaZ signifikant schlechter ab ($Wald\text{-}\chi^2(1) = 13.40; p < .01$).

5.4 Latente Mehrgruppenregressionsanalyse und latenter Mittelwertvergleich der Lesekompetenzwerte: Gesamtmodell

Das Gesamtmodell als Mehrgruppenregressionsanalyse mit den latenten Variablen phonologische Bewusstheit, prosodische Sensitivität, Wortschatz und Grammatikkompetenz als Prädiktoren der Lesekompetenz zeigt für die monolinguale Gruppe bedeutsame Beiträge der Grammatikkompetenz ($\beta = .35; p = .03$) und des Wortschatzes auf ($\beta = .33; p = .03$).

Für die Lesekompetenz der mehrsprachigen Gruppe mit DaE erweist sich der Wortschatz als signifikanter Prädiktor ($\beta = .51; p = .04$), der Pfadkoeffizient der Grammatikkompetenz verfehlt knapp die statistische Signifikanz ($\beta = .27; p = .05$).

In der Gruppe der mehrsprachigen Jugendlichen mit DaZ erweist sich nur die Grammatikkompetenz als bedeutsamer Prädiktor der Lesekompetenz ($\beta = .31; p = .00$). In keiner der drei Gruppen leisten prosodische Sensitivität oder phonologische Bewusstheit bei gleichzeitiger Berücksichtigung von Wortschatz und Grammatikkompetenz signifikante Beiträge zur Aufklärung der Leseleistungen (s. Tab. 5).

Modell 1			
	Monolingual (n = 107)	Mehrsprachig/DaE (n = 105)	Mehrsprachig/DaZ (n = 198)
Pfadkoeffizienten	β (SE)	β (SE)	β (SE)
Phonologie	.11 (.15)	.07 (.16)	-.02 (.12)
Prosodie	.09 (.13)	.05 (.12)	.17 (.11)
Wortschatz	.35* (.16)	.51* (.25)	.18 (.12)
Grammatik	.33* (.16)	.27 (.14)	.31* (.10)
Intercept	α (SE)	α (SE)	α (SE)
	0.00 _a	-0.14 _a (0.39)	-.11 _a (0.31)

Anmerkungen: $\chi^2/df = 418.58/548, p > .99$; RMSEA < .01; CFI > .99; TLI > .99. Gerichtete Pfade der manifesten Kontrollvariablen Geschlecht, Intelligenz sowie sozioökonomischer-soziokultureller Hintergrund auf alle latenten Variablen. Der Mittelwert der monolingualen Probanden wurde zwecks Skalierung auf Null fixiert. Latente Mittelwerte mit ungleichen Fußindizes unterscheiden sich mit $p < .05$ statistisch überzufällig. * $p < .05$.

Tab. 5: Latente Mehrgruppenregressionsanalyse und latenter Mittelwertvergleich mit Lesekompetenz als abhängiger Variable. Gesamtmodell mit den latenten Prädiktoren phonologische Bewusstheit, prosodische Sensitivität, Wortschatz und Grammatikkompetenz (Modell 1)

Die Gruppenunterschiede im Lesen zwischen den monolingualen Jugendlichen und den beiden Gruppen mehrsprachiger Jugendlicher sind im Gesamtmodell nicht signifikant (monolingual vs. DaE: $Wald\text{-}\chi^2(1) = 0.12$; $p = .73$; monolingual vs. DaZ: $Wald\text{-}\chi^2(1) = 0.13$; $p = .72$). Die Gruppen der mehrsprachigen Jugendlichen unterscheiden sich auch nicht untereinander ($Wald\text{-}\chi^2(1) = 0.01$; $p = .94$).

5.5 Latente Mehrgruppenregressionsanalyse und latenter Mittelwertvergleich der Lesekompetenz: Die Bedeutung des Wortschatzes

Bei Ausschluss der Grammatikkompetenz aus dem Gesamtmodell (s. Tab. 6, Modell 2) erweist sich in allen drei Gruppen ausschließlich der Wortschatz als signifikanter Prädiktor der Lesekompetenz (monolingual: $\beta = .42$; $p = .01$; DaE: $\beta = .57$; $p = .03$; DaZ: $\beta = .28$; $p = .02$). Es finden sich weder zwischen den monolingualen Jugendlichen und den beiden mehrsprachigen Gruppen (monolingual vs. DaE: $Wald\text{-}\chi^2(1) = 1.36$; $p = .24$; monolingual vs. DaZ: $Wald\text{-}\chi^2(1) = 1.73$; $p = .19$) noch zwischen den beiden mehrsprachigen Gruppen latente Mittelwertunterschiede im Lesen (DaE vs. DaZ: $Wald\text{-}\chi^2(1) = .11$; $p = .74$).

Modell 2			
	Monolingual (n = 107)	Mehrsprachig/DaE (n = 105)	Mehrsprachig/DaZ (n = 198)
Pfadkoeffizienten	β (SE)	β (SE)	β (SE)
Phonologie	.23 (.14)	.12 (.16)	.10 (.11)
Prosodie	.12 (.14)	.11 (.12)	.20 (.11)
Wortschatz	.42* (.15)	.57* (.27)	.28* (.12)
Grammatik			
Intercept	α (SE)	α (SE)	α (SE)
	0.00 _a	−0.44 _a (0.38)	−.33 _a (0.31)

Anmerkungen: $\chi^2/df = 191.33/265$, $p > .99$; RMSEA < .01; CFI > .99; TLI > .99. Gerichtete Pfade der manifesten Kontrollvariablen Geschlecht, Intelligenz sowie sozioökonomischer-soziokultureller Hintergrund auf alle latenten Variablen. Der Mittelwert der monolingualen Probanden wurde zwecks Skalierung auf Null fixiert. Latente Mittelwerte mit ungleichen Fußindizes unterscheiden sich mit $p < .05$ statistisch überzufällig. * $p < .05$.

Tab. 6: Latente Mehrgruppenregressionsanalyse und latenter Mittelwertvergleich mit Lesekompetenz als abhängiger Variable. Ausschluss der Grammatikkompetenz aus dem Gesamtmodell (Modell 2)

5.6 Latente Mehrgruppenregressionsanalyse und latenter Mittelwertvergleich der Lesekompetenz: Die Bedeutung von phonologischer Bewusstheit und prosodischer Sensitivität

Mit phonologischer Bewusstheit und prosodischer Sensitivität als Prädiktoren (vgl. Tab. 7, Modell 3), d.h. unter Ausschluss der Grammatikkompetenz und des Wortschatzes aus dem Gesamtmodell, erweist sich die prosodische Sensitivität als bedeutsam für die Lesekompetenzwerte der mehrsprachigen Jugendlichen mit DaZ ($\beta = .23$; $p = .04$).

Unter alleiniger Kontrolle der phonologischen und prosodischen Fähigkeiten ergeben sich Unterschiede in der Lesekompetenz zwischen monolingualen Jugendlichen und mehrsprachigen Jugendlichen mit DaE ($\text{Wald-}\chi^2(1) = 4.82$; $p = .03$) sowie zwischen monolingualen Jugendlichen und mehrsprachigen Jugendlichen mit DaZ ($\text{Wald-}\chi^2(1) = 4.16$; $p = .04$). Es zeigen sich jedoch keine statistisch bedeutsamen Unterschiede im Lesen zwischen den Gruppen mehrsprachiger Jugendlicher ($\text{Wald-}\chi^2(1) = 0.28$; $p = .60$).

Modell 3			
	Monolingual ($n = 107$)	Mehrsprachig/DaE ($n = 105$)	Mehrsprachig/DaZ ($n = 198$)
Pfadkoeffizienten	β (SE)	β (SE)	β (SE)
Phonologie	.18 (.16)	.15 (.16)	.13 (.10)
Prosodie	.16 (.14)	.13 (.13)	.23* (.11)
Wortschatz			
Grammatik			
Intercept	α (SE)	α (SE)	α (SE)
	0.00 _a	-0.68 _b (0.29)	-0.57 _b (0.28)

Anmerkungen: $\chi^2/\text{df} = 117.19/166$, $p > .99$; RMSEA < .01; CFI > .99; TLI > .99. Gerichtete Pfade der manifesten Kontrollvariablen Geschlecht, Intelligenz sowie sozioökonomischer-soziokultureller Hintergrund auf alle latenten Variablen. Der Mittelwert der monolingualen Probanden wurde zwecks Skalierung auf Null fixiert. Latente Mittelwerte mit ungleichen Fußindizes unterscheiden sich mit $p < .05$ statistisch überzufällig. * $p < .05$.

Tab. 7: Latente Mehrgruppenregressionsanalyse und latenter Mittelwertvergleich mit Lesekompetenz als abhängiger Variable sowie phonologischer Bewusstheit und prosodischer Sensitivität als Prädiktoren (Modell 3)

6. Diskussion

6.1 Vergleich der latenten Mittelwerte

Das Muster großer Mittelwertunterschiede, wie es sich in der DESI-Studie (Hesse et al., 2008) gezeigt hatte, konnte in unserer Untersuchung im unteren Leistungsbereich nicht in diesem Umfang bestätigt werden. Insbesondere im Bereich der Lesekompetenz fanden sich in dieser selektierten Stichprobe keine signifikanten Unterschiede zwischen den beiden mehrsprachigen Gruppen. Im Vergleich zur monolingualen Gruppe schnitten beide mehrsprachigen Gruppen signifikant schlechter ab. Im Hinblick auf den Wortschatz fanden sich wie erwartet substanzielle Nachteile der mehrsprachigen Jugendlichen mit DaZ gegenüber mehrsprachigen Jugendlichen mit DaE. Die mehrsprachigen Jugendlichen mit DaZ verfügten auch erwartungskonform über einen signifikant geringeren Wortschatz als monolinguale Jugendliche. Im Sinne von Cummins (1981) lässt sich dieser Befund als Folge des kürzeren und weniger umfangreichen Sprachkontakts interpretieren. Überraschenderweise finden sich bei mehrsprachigen Jugendlichen mit Deutsch als Erstsprache keine signifikant schlechteren Wortschatzwerte als bei monolingualen Jugendlichen. Dies deckt sich mit Befunden, wonach sich Rückstände im Wortschatz der Verkehrssprache aufgrund von Mehrsprachigkeit ab einem gewissen Sprachentwicklungsstand nicht mehr anhand der absoluten Größe des Wortschatzes manifestieren, sondern vielmehr anhand der Schnelligkeit des lexikalischen Abrufs (Bialystok, 2010).

Im Hinblick auf die formalen Bereiche der Sprache fanden sich bedeutsame Fähigkeitsunterschiede zwischen den mehrsprachigen Probanden im Bereich der prosodischen Sensitivität. Mehrsprachige Probanden mit DaZ schnitten in diesem Bereich signifikant schlechter ab als mehrsprachige Probanden mit DaE. Da die mehrsprachige Gruppe mit DaE hierbei tendenziell sogar bessere Werte erzielte als die monolinguale Gruppe, bleibt jedoch offen, ob die Unterschiede zwischen den mehrsprachigen Gruppen als Beleg für eine spezifische Schwäche der DaZ-Gruppe gedeutet werden dürfen oder ob die Resultate vielmehr auf eine besondere Stärke der DaE-Gruppe in diesem Bereich verweisen.

Im Bereich der phonologischen Bewusstheit ließen sich keine latenten Mittelwertunterschiede nachweisen. Die wortschatzneutrale Erfassung der phonologischen Bewusstheit mithilfe von Pseudowörtern mag diesen Befund zum Teil erklären. Schwierigkeiten der Lautdiskrimination im Zuge des Zweitspracherwerbs können jedoch ohnehin nur dann auftreten, wenn die fokussierten phonetischen Kategorien der Zweitsprache nicht ebenfalls in der Erstsprache existieren (vgl. Kuhl, 2004). Die verwendeten Lautkategorisierungsaufgaben wurden jedoch nicht im Hinblick darauf konstruiert, möglichst stark mit den Phoneminventaren anderer Sprachen, z.B. dem des Türkischen, zu kontrastieren. Eine detailliertere Auswertung unter Berücksichtigung des phonologischen Repertoires der konkurrierenden nicht deutschen Erstsprachen könnte die Einschätzung von Kompetenzunterschieden vielleicht noch präzisieren.

Hinsichtlich der Grammatikkompetenz, der dritten der rein formal-sprachlichen Domänen, erzielten beide multilingualen Gruppen wesentlich geringere Werte als die Gruppe der monolingualen Jugendlichen. Beide mehrsprachigen Gruppen unterschieden sich jedoch nicht signifikant voneinander. Die geringen Unterschiede zwischen den beiden mehrsprachigen Gruppen können allerdings nicht auf unerwartet starke Leistungen der mehrsprachigen Jugendlichen mit DaZ zurückgeführt werden. Vielmehr scheint die Grammatikkompetenz der mehrsprachigen Jugendlichen mit DaE trotz relativ hoher Wortschatzwerte überraschend gering zu sein. Dies könnte darauf hinweisen, dass der simultane Erwerb zweier linguistischer Systeme in dieser relativ leistungsschwachen Stichprobe zu Interferenzen geführt hat.

6.2 Aufklärung von Lesekompetenzunterschieden

Für die Lesekompetenz von mehrsprachigen Jugendlichen mit DaZ ist prosodische Sensitivität ein bedeutsamer Prädiktor. Die regressionsanalytische Kontrolle der prosodischen Sensitivität genügt jedoch nicht, um die Lesekompetenzrückstände gegenüber den monolingualen Schülern zu erklären. Phonologische Bewusstheit kann nicht mit den Rückständen der mehrsprachigen Schülerinnen und Schüler mit DaZ im Bereich der Lesekompetenz in Zusammenhang gebracht werden. Grammatikkompetenz erwies sich im Gesamtmodell als einziger bedeutsamer Faktor zur Erklärung der Lesekompetenzrückstände der mehrsprachigen Gruppe mit DaZ. Im Regressionsmodell ohne Einbezug der Grammatikkompetenz erwies sich jedoch auch der Wortschatz als bedeutsamer Prädiktor der Lesekompetenz, der die Rückstände der mehrsprachigen Gruppe mit DaZ erklären kann. Die Nachteile von mehrsprachigen Jugendlichen mit DaZ im Lesen können offensichtlich durch die geteilten Varianzanteile von Wortschatz und Grammatikkompetenz erklärt werden. Die Erkenntnis, dass Wortschatz und Grammatikkompetenz bei Jugendlichen mit DaZ in hohem Umfang äquivalente Funktionen übernehmen, deckt sich mit der Annahme, dass bei Lernern einer Zweitsprache Grammatikkompetenz und Wortschatz gleichermaßen vom deklarativen Gedächtnis abhängen, wohingegen die Grammatik bei Muttersprachlern auf dem prozeduralen Gedächtnis aufbaut (Ullman, 2001). Im Sinne eines emergenztheoretischen Modells des Grammatikerwerbs (Bates & Goodman, 1998) könnte man auch annehmen, dass bei DaZ-Jugendlichen infolge eines zu geringen Wortschatzes noch keine differenzierte Modularisierung von Wortschatz und Grammatik stattgefunden hat. Darüber hinaus erscheint es ebenfalls möglich, dass der Wortschatz der Jugendlichen mit DaZ so gering ausfällt, dass er nicht über die basalen Kenntnisse hinausgeht, die auch innerhalb des Grammatiktests benötigt werden. Eine zusätzliche Varianzaufklärung seitens des Wortschatzes könnte auch auf diese Weise unterbunden sein.

Für mehrsprachige Jugendliche mit DaE ist nur der Wortschatz ein bedeutsamer Prädiktor der Lesekompetenz. Die regressionsanalytische Kontrolle des Wortschatzes nivelliert auch in dieser Gruppe die Lesekompetenzrückstände gegenüber der monolingualen Gruppe. Der Vorhersagewert der Grammatikkompetenz für die Lesekompetenz

ist unter gleichzeitigem Einbezug des Wortschatzes nicht signifikant. Das knappe Verfehlen des Fünfprozentniveaus bei einer vergleichsweise kleinen Stichprobe legt jedoch nahe, dass die funktionale Überschneidung von Wortschatz und Grammatikkompetenz in dieser Gruppe nicht so stark ausfällt wie in der mehrsprachigen Gruppe mit DaZ.

Für den Leseerfolg der monolingualen Gruppe erweisen sich sowohl Grammatikkompetenz als auch Wortschatz als prädiktive Faktoren.

6.3 Zusammenfassung

Die vorliegende Studie mit leseschwachen Schülerinnen und Schülern legt nahe, dass die Spracherwerbsreihenfolge für die Leseschwierigkeiten von mehrsprachigen Schülerinnen und Schülern eine verhältnismäßig geringe Bedeutung einnimmt. Mehrsprachige Jugendliche mit DaZ und DaE schneiden im Vergleich zu der monolingualen Gruppe gleichermaßen schlecht ab. Die Rückstände im Bereich des Leseverstehens von mehrsprachigen Jugendlichen werden unabhängig vom Zeitpunkt des Erwerbsbeginns der deutschen Sprache primär mit eingeschränkten Fähigkeiten im Überschneidungsbereich von Grammatikkompetenz und Wortschatz erklärt.

Bei aller theoretisch begründeten Plausibilität kann die vorliegende querschnittliche quasiexperimentelle Untersuchung keine Aussagen über kausale Mechanismen treffen. Darüber hinaus geht die Fokussierung der Studie auf Schülerinnen und Schüler in den unteren Leistungsbereichen zwangsläufig mit einer gewissen künstlichen Einschränkung der Leistungsvarianz in einzelnen Gruppen einher und schränkt dadurch die Generalisierbarkeit der Mittelwertvergleiche ein. Im Rahmen der vorliegenden Gruppenvergleiche kann auch nicht ausgeschlossen werden, dass das Ausbleiben der vorhergesagten Fähigkeitsunterschiede zwischen den mehrsprachigen Gruppen in den Bereichen Lesekompetenz, Grammatikkompetenz oder phonologische Bewusstheit auf unterschiedliche ethnische Kompositionen oder familiäre Sprachgewohnheiten zurückzuführen ist. Zuletzt muss einschränkend bemerkt werden, dass nicht mit Sicherheit festgestellt werden kann, ob die retrospektive Erfassung der Erstsprache valide den chronologischen Verlauf des Spracherwerbs abbildet. Es mag sein, dass die Sprachdominanz, d.h. die relative Nutzungshäufigkeit der Sprachen oder auch das ethnische Identitätsempfinden, als Anhaltspunkt bei autobiografischen Unsicherheiten gedient hat. Die Inkonsistenzen bei der exakten zeitlichen Beschreibung der Sprachbiografien unterstützen diese Vermutung. Darüber hinaus scheint die mangelnde Sensitivität für prosodisch markierte Phrasengrenzen in einer Zweitsprache, wie sie sich auch in dieser Studie andeutet, eher von der Dominanz einer konkurrierenden Sprache abzuhängen, in der andere Segmentationshinweise (z.B. Vokalharmonien) bedeutsam sind, und in geringerem Ausmaß von der Erwerbsreihenfolge bzw. dem Zeitpunkt des Erwerbsbeginns (Cutler, 2001).

Diesen Einschränkungen zum Trotz vermitteln die vorliegenden Befunde detaillierte Hinweise auf allgemeine und differenzielle Bedingungsfaktoren der Leseschwierigkeiten von mehrsprachigen Leserinnen und Lesern mit DaE und DaZ.

Literatur

- Artelt, C., McElvany, N., Christmann, U., Richter, T., Groeben, N., Köster, J., Schneider, W., Stanat, P., Ostermeier, C., Schiefele, U., Valtin, R., & Ring, K. (2007). *Förderung von Lesekompetenz – Expertise* (1. Aufl./unveränderter Nachdruck). Berlin: Bundesministerium für Bildung und Forschung.
- Asparouhov, T., & Muthén, B. O. (2010). *Chi-square statistics with multiple imputation. Technical Appendix, Version 2*. <http://www.statmodel.com/download/MI7.pdf> [27.05.2011].
- Bates, E., & Goodman, J. (1998). On the emergence of grammar from the lexicon. In B. MacWhinney (Hrsg.), *Emergence of Language* (S. 29-81). Hillsdale: Erlbaum.
- Bialystok, E. (2010). Bilingualism. *Wiley Interdisciplinary Reviews: Cognitive Science*, 1(4), 559-572.
- Bialystok, E., & Hakuta, K. (1999). Confounded age: Linguistic and cognitive factors in age differences for second language acquisition. In D. P. Birdsong (Hrsg.), *Second language acquisition and the critical period hypothesis* (S. 161-181). Hillsdale: Erlbaum.
- Bialystok, E., Majumder, S., & Martin, M. (2003). Developing phonological awareness: Is there a bilingual advantage? *Applied Psycholinguistics*, 24(1), 27-44.
- Buse, A. (1982). The likelihood ratio, wald, and lagrange multiplier test: An expository note. *The American Statistician*, 36(3), 153-157.
- Chiappe, P., Siegel, L. S., & Wade-Woolley, L. (2002). Linguistic diversity and the development of reading skills: A longitudinal study. *Scientific Studies of Reading*, 6(4), 369-400.
- Cummins, J. (1981). Age on arrival and immigrant second language learning in Canada. *Applied Linguistics*, 11(2), 132-149.
- Cutler, A. (2001). Listening to a second language through the ears of a first. *Interpreting*, 5(1), 1-23.
- Ehmke, T., & Siegle, T. (2005). ISEI, ISCED, HOMEPOS, ESCS. Indikatoren der sozialen Herkunft bei der Quantifizierung von sozialen Disparitäten. *Zeitschrift für Erziehungswissenschaft*, 8(4), 521-540.
- Gustafsson, J. E., & Balke, G. (1993). General and specific abilities as predictors of school achievement. *Multivariate Behavioral Research*, 28(4), 407-434.
- Hancock, G. (2001). Effect size, power, and sample size determination for structured means modeling and mimic approaches to between-groups hypothesis testing of means on a single latent construct. *Psychometrika*, 66(3), 373-388.
- Hancock, G., & Mueller, R. (2006). *Structural equation modeling*. Greenwich: Information Age Publishing.
- Heller, K. A., & Perleth, C. (2000). *Kognitiver Fähigkeitstest für 4. bis 12. Klassen, Revision: KFT 4-12+R*. Weinheim: Beltz.
- Hesse, H.-G., Göbel, K., & Hartig, J. (2008). Sprachliche Kompetenzen von mehrsprachigen Jugendlichen und Jugendlichen nicht-deutscher Erstsprache. In DESI-Konsortium (Hrsg.), *Unterricht und Kompetenzerwerb in Deutsch und Englisch* (S. 208-230). Weinheim: Beltz.
- Kim, S., & Hagtvet, K. (2003). The impact of misspecified item parceling on representing latent variables in covariance structure modeling: A simulation study. *Structural Equation Modeling*, 10(1), 101-127.
- Kovelman, I., Baker, S. A., & Petitto, L. A. (2008). Age of first bilingual language exposure as a new window into bilingual reading development. *Bilingualism: Language and Cognition*, 11(2), 203-223.
- Kuhl, P. K. (2004). Early language acquisition: Cracking the speech code. *Nature Reviews Neuroscience*, 5, 831-843.
- Kuo, L.-J., & Anderson, R. (2008). Conceptual and methodological issues in comparing metalinguistic awareness across languages. In K. Koda & A. Zehler (Hrsg.), *Learning to read across languages* (S. 39-46). New York: Routledge.

- Limbird, C., & Stanat, P. (2006). Prädiktoren von Leseverständnis bei Kindern deutscher und türkischer Herkunftssprache: Ergebnisse einer Längsschnittstudie. In A. Ittel & H. Merckens (Hrsg.), *Veränderungsmessung und Längsschnittstudien in der empirischen Erziehungswissenschaft* (S. 93-123). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Lüdtke, O., Robitzsch, A., Trautwein, U., & Köller, O. (2007). Umgang mit fehlenden Werten in der psychologischen Forschung. Probleme und Lösungen. *Psychologische Rundschau*, 58(2), 103-117.
- Muthén, L. K., & Muthén, B. O. (1998-2010). *Mplus Users' Guide* (6. Aufl.). Los Angeles: Muthén & Muthén.
- Naumann, J., Artelt, C., Schneider, W., & Stanat, P. (2010). Lesekompetenz von PISA 2000 bis PISA 2009. In E. Klieme, C. Artelt, J. Hartig, N. Jude, O. Köller, M. Prenzel, W. Schneider & P. Stanat (Hrsg.), *PISA 2009. Bilanz nach einem Jahrzehnt* (S. 23-71). Münster: Waxmann.
- OECD (2009). *PISA data analysis manual: SPSS* (2. Aufl.). Paris: OECD.
- Penner, Z. (2005). *Auf dem Weg zur Sprachkompetenz. Neue Perspektiven der sprachlichen Frühförderung bei Migrantenkindern. Ein Arbeitsbuch*. Frauenfeld: Kon-Lab GmbH.
- Ramm, G., Prenzel, M., Baumert, J., Blum, W., Lehmann R., Leutner, D., Neubrand, M., Pekrun, R., Rolff, H.-G., Rost, J., & Schiefele, U. (2006). *PISA 2003. Dokumentation der Erhebungsinstrumente*. Münster: Waxmann.
- Saffran, J. R., Aslin, R., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, 274(5294), 1926-1928.
- Segerer, R., Marx, A., & Marx, P. (im Druck). Unlösbare Items im KFT4-12+R. *Diagnostica* [im Druck].
- Stanat, P., & Christensen, G. (2006). *Schulerfolg von Jugendlichen mit Migrationshintergrund im internationalen Vergleich. Eine Analyse von Voraussetzungen und Erträgen schulischen Lernens im Rahmen von PISA 2003* (Bildungsforschung, Bd. 19). Berlin: Bundesministerium für Bildung und Forschung.
- Stanat, P., Rauch, D., & Segeritz, M. (2010). Schülerinnen und Schüler mit Migrationshintergrund. In E. Klieme, C. Artelt, J. Hartig, N. Jude, O. Köller, M. Prenzel, W. Schneider & P. Stanat (Hrsg.), *PISA 2009. Bilanz nach einem Jahrzehnt* (S. 200-230). Münster: Waxmann.
- Stock, C., & Schneider, W. (in Vorb.). *BAKO 5-9: Basiskompetenzen für Lese-Rechtschreibleistungen*. Göttingen: Hogrefe.
- Tiedemann, J., & Billmann-Mahecha, E. (2007). Leseverständnis, Familiensprache und Freizeitsprache. Ergebnisse aus der Hannoverschen Grundschulstudie. *Zeitschrift für Pädagogische Psychologie*, 21(1), 41-49.
- Ullman, M. (2001). The neural basis of lexicon and grammar in first and second language: the declarative/procedural model. *Bilingualism: Language and Cognition*, 4(2), 105-122.
- Vandenberg, R., & Lance, C. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, 3(1), 4-69.
- Verhoeven, L. (2000). Components in early second language reading and spelling. *Scientific Studies of Reading*, 4(4), 313-330.
- von Davier, M., Gonzalez, E., & Mislevy, R. (2009). What are plausible values and why are they useful? *IERI Monograph Series*, 2, 9-36.
- Walter, O. (2008). Herkunftsassoziierte Disparitäten im Lesen, in der Mathematik und in den Naturwissenschaften: Ein Vergleich zwischen PISA 2000, PISA 2003 und PISA 2006. *Zeitschrift für Erziehungswissenschaft, Sonderheft 10*, 149-168.
- Wartenburger, I., Heekeren, H., Abutalebi, J., Cappa, S., Villringer, A., & Perani, D. (2003). Early setting of grammatical processing in the bilingual brain. *Neuron*, 37(1), 159-170.
- Weber, J., Marx, P., & Schneider, W. (2007). Die Prävention von Lese-Rechtschreibschwierigkeiten bei Kindern mit nichtdeutscher Herkunftssprache durch ein Training der phonologischen Bewusstheit. *Zeitschrift für Pädagogische Psychologie*, 21(1), 65-75.

- Weber-Fox, C., & Neville, H. (1996). Maturational constraints on functional specializations for language processing: ERP and behavioral evidence in bilingual speakers. *Journal of Cognitive Neuroscience*, 8(3), 231-256.
- Weiß, R. (2008). *CFT 20-R. Grundintelligenztest Skala 2 – Revision*. Göttingen: Hogrefe.
- Wu, M. (2005). The role of plausible values in large-scale surveys. *Studies in Educational Evaluation*, 31(2-3), 114-128.

Anschrift der Autor(inn)en

Dipl.-Psych. Robin Segerer, Julius-Maximilians-Universität Würzburg, Lehrstuhl für Psychologie IV, Röntgenring 10, 97070 Würzburg, Deutschland
E-Mail: segerer@psychologie.uni-wuerzburg.de

Dipl.-Psych. Alexandra Marx, Humboldt-Universität zu Berlin, Institut zur Qualitätsentwicklung im Bildungswesen (IQB), Unter den Linden 6, 10099 Berlin, Deutschland
E-Mail: alexandra.marx@iqb.hu-berlin.de

Prof. Dr. Petra Stanat, Humboldt-Universität zu Berlin, Institut zur Qualitätsentwicklung im Bildungswesen (IQB), Unter den Linden 6, 10099 Berlin, Deutschland
E-Mail: petra.stanat@iqb.hu-berlin.de

Prof. Dr. Wolfgang Schneider, Julius-Maximilians-Universität Würzburg, Lehrstuhl für Psychologie IV, Wittelsbacherplatz 1, 97074 Würzburg, Deutschland
E-Mail: schneider@psychologie.uni-wuerzburg.de

Dr. Thorsten Roick, Humboldt-Universität zu Berlin, Abteilung für Erziehungswissenschaftliche Methodenlehre, Geschwister-Scholl-Str. 7, 10099 Berlin, Deutschland
E-Mail: thorsten.roick@hu-berlin.de

Dr. Peter Marx, Julius-Maximilians-Universität Würzburg, Lehrstuhl für Psychologie IV, Wittelsbacherplatz 1, 97074 Würzburg, Deutschland
E-Mail: marx@psychologie.uni-wuerzburg.de

Trends und Veränderungen

Timo Ehmke/Eckhard Klieme/Petra Stanat

Veränderungen der Lesekompetenz von PISA 2000 nach PISA 2009

Die Rolle von Unterschieden in den Bildungswegen und in der Zusammensetzung der Schülerschaft

1. Einleitung

Studien wie das *Programme for International Student Assessment* (PISA) zielen im Sinne des Bildungsmonitoring primär darauf ab, den teilnehmenden Staaten Informationen über die Erträge ihrer schulischen Systeme zu liefern. In Kernbereichen schulischer Bildung werden die von Schülerinnen und Schülern erreichten Kompetenzen erfasst, um Hinweise auf Stärken und Schwächen zu erhalten und um möglichen Handlungsbedarf zu identifizieren. Der zyklische Charakter von PISA, der regelmäßige Querschnitterhebungen in Abständen von drei Jahren vorsieht, soll es zudem ermöglichen, Veränderungen über die Zeit zu bestimmen. Eine bislang kaum eingelöste Intention solcher Trend-Beobachtungen ist zudem, für die einzelnen Staaten oder im internationalen Vergleich mögliche Ursachen für langfristige systemische Entwicklungen zu identifizieren. Vor allem Bildungsökonominnen (vgl. Hanushek & Wössmann, 2010), aber auch die Statistiker der OECD (vgl. etwa OECD, 2010) bemühen sich, solche Ursachen im Bereich der sozioökonomischen Rahmenbedingungen und im Bereich der Bildungspolitik aufzuspüren, um zum Beispiel mögliche Effekte von Reformbemühungen abzuschätzen.

Das in PISA umgesetzte Forschungsdesign sieht für die Erhebungen wechselnde Schwerpunkte vor. Für den jeweiligen Schwerpunktbereich steht eine längere Testzeit zur Verfügung, so dass er mit einer größeren Anzahl von Aufgaben breiter erfasst werden kann. Im Jahr 2000 war der Schwerpunkt die Lesekompetenz, im Jahr 2003 die mathematische Grundbildung und im Jahr 2006 die naturwissenschaftliche Grundbildung. Die anderen Kompetenzbereiche bildeten zu diesen Erhebungszeitpunkten jeweils die NebenkompONENTEN. Obwohl auch für die NebenkompONENTEN Aussagen über Veränderungen getroffen werden können, sind die Schätzungen deutlich verlässlicher, wenn eine Kompetenzdomäne zu beiden betrachteten Messzeitpunkten den Schwerpunkt bildete. Dies war erstmals 2009 der Fall, als die Lesekompetenz nach 2000 zum zweiten Mal als Schwerpunktbereich erfasst wurde. Im 2010 veröffentlichten PISA-Bericht

(Klieme, Artelt et al., 2010) konnten also verlässliche Aussagen darüber getroffen werden, inwieweit die von Schülerinnen und Schülern in Deutschland erreichte Lesekompetenz über einen Zeitraum von neun Jahren angestiegen ist und sich die mit der sozialen Herkunft und dem Migrationshintergrund verbundenen Disparitäten, die 2000 in Deutschland besonders ausgeprägt waren, reduziert haben. Diese Befunde werden im Folgenden kurz zusammengefasst, um anschließend auf mögliche Ursachen für die beobachteten Veränderungen einzugehen.

1.1 *Veränderungen in der Lesekompetenz zwischen PISA 2000 und PISA 2009*

Im Dezember 2010 wurden die Ergebnisse der vierten Erhebung im Rahmen des *Programme for International Student Assessment* (OECD, 2009) veröffentlicht. Damit liegen für vier über einen Zeitraum von neun Jahren verteilte Messzeitpunkte repräsentative Querschnittsinformationen über Kompetenzen von Schülerinnen und Schülern, zentrale Hintergrundmerkmale sowie Schul- und Unterrichtsbedingungen vor. Die über sogenannte Linking-Designs (Holland & Dorans, 2006) verknüpften Kompetenzmessungen und die Erfassung von vergleichbaren Herkunftsmerkmalen der Schülerinnen und Schüler ermöglichen Trendaussagen darüber, wie sich wichtige Kennwerte des Bildungssystems über die Zeit entwickelt haben. In den entsprechenden Analysen konnten zwei wesentliche Verbesserungen im Verlaufe des letzten Jahrzehnts festgestellt werden (Klieme, Jude, Baumert & Prenzel, 2010): (1) ein bedeutsamer Anstieg in der Lesekompetenz der Jugendlichen und (2) eine Abnahme der sozialen und migrationsbezogenen Disparitäten.

Ein Anstieg der Lesekompetenz bei den Jugendlichen in Deutschland war nach PISA 2000 auch schon in den Ergebnissen für 2003 und 2006 zu beobachten. Dieser ließ sich jedoch nicht inferenzstatistisch absichern. In PISA 2009 lag die Lesekompetenz schließlich mit 497 Punkten statistisch signifikant über dem im Jahr 2000 gemessenen Wert von 484 Punkten (Naumann, Artelt, Schneider & Stanat, 2010, S. 59). Ein Vergleich zwischen den prozentualen Anteilen von Schülerinnen und Schülern auf den Kompetenzstufen im Lesen ergab, dass sich vor allem die Anteile auf den untersten Kompetenzstufen über die Zeit verringert hatten (Naumann et al., 2010, S. 62).

Mit dem Anstieg der Lesekompetenz im unteren Leistungsbereich haben sich zwischen PISA 2000 und PISA 2009 gleichzeitig auch die sozialen Disparitäten der Lesekompetenz verringert. So reduzierte sich die Steigung des sozialen Gradienten in der Lesekompetenz – ein für den internationalen Vergleich wichtiger Kennwert sozialer Disparitäten – in Deutschland statistisch bedeutsam von 45 Punkten in PISA 2000 auf 35 Punkte in PISA 2009 (Ehmke & Jude, 2010, S. 241). Damit waren für Deutschland zwar immer noch ausgeprägte soziale Disparitäten im Kompetenzerwerb festzustellen, es deutete sich jedoch eine Entwicklung hin zu mehr Chancengleichheit im Bildungssystem an.

Für die Lesekompetenz von Jugendlichen mit Migrationshintergrund ergab der internationale Vergleich zwei zentrale Befunde. Erstens zeigte sich, dass die Kompetenznachteile von Jugendlichen mit Migrationshintergrund in Deutschland, wie auch in mittel- und nordeuropäischen Vergleichsstaaten, sowohl in der ersten Generation (Eltern und Jugendlicher im Ausland geboren) als auch in der zweiten Generation (Eltern im Ausland geboren, Jugendlicher in Deutschland geboren) weiterhin groß sind. Zweitens konnten sich aber die Jugendlichen aus zugewanderten Familien seit PISA 2000 im Lesen signifikant verbessern. Dieser positive Trend fiel in der ersten Generation größer aus (+33 Punkte) als in der zweiten Generation (+24 Punkte) und ließ sich auch nur für die erste Generation gegen den Zufall absichern (Stanat, Rauch & Segeritz, 2010, S. 225). Da gleichzeitig für die Jugendlichen ohne Migrationshintergrund keine signifikanten Kompetenzsteigerungen zu verzeichnen waren, haben sich die migrationsbezogenen Disparitäten reduziert.

Insgesamt ergab sich für die Gruppe der Schülerinnen und Schüler mit Migrationshintergrund ein durchschnittlicher Zuwachs zwischen 2000 und 2009 von 26 Testpunkten (Stanat et al., 2010, S. 212). Berücksichtigt man, dass inzwischen etwa jeder vierte Jugendliche der PISA-Population in Deutschland einen Migrationshintergrund aufweist, bedeutet dies, dass die Kompetenzzuwächse dieser Gruppe rechnerisch knapp 7 der 13 Punkte des gesamten Trendgewinns auf nationaler Ebene ausmachen. Mit anderen Worten: Etwa die Hälfte des Zuwachses in der Lesekompetenz zwischen 2000 und 2009 ist erklärbar durch den Leistungsanstieg bei Jugendlichen mit Migrationshintergrund. Diese Abschätzung gibt einen ersten wichtigen Hinweis auf mögliche Ursachen des beobachteten Trends.

1.2 *Mögliche Ursachen von Veränderungen zwischen PISA 2000 und PISA 2009*

Studien, die primär dem Bildungsmonitoring dienen, lassen in der Regel keine gesicherten Aussagen darüber zu, worauf die beobachteten Veränderungen zurückzuführen sind. Implizit wird häufig davon ausgegangen, dass Reformbemühungen, die zwischen den beiden Messzeitpunkten unternommen wurden, für die beobachteten Verbesserungen verantwortlich sind bzw. dass die Bemühungen bei ausbleibenden Verbesserungen nicht effektiv waren. Inwieweit diese Annahmen zutreffen, kann anhand der Daten allerdings kaum bestimmt werden. Dies liegt an der Vermischung („Konfundierung“) unterschiedlicher Effekte (z.B. Veränderungen des sozioökonomischen und kulturellen Kontextes, bildungspolitische Struktur- und Ressourcenentscheidungen, pädagogische Qualitätsentwicklung), deren spezifische Wirkungen sich – zumal bei einer geringen „Fall“-Zahl von 30 bis 70 Staaten – nur schwer quantifizieren und schon gar nicht kausal deuten lassen. Entsprechende Versuche beschränken sich daher in der Regel darauf, dass Veränderungen „mehr oder weniger plausibel nachvollzogen werden, indem man kulturelle, soziale und bildungspolitische Entwicklungen resümiert“ (vgl. Klieme, Jude et al., 2010, S. 292; dort finden sich entsprechende Interpretationsver-

suche zu Staaten mit insgesamt hohen PISA-Ergebnissen oder auffallend positiven Veränderungen).

Bezogen auf Deutschland wird vielfach aus bildungspolitischer und auch aus pädagogischer Sicht angenommen oder zumindest erhofft, dass sich die zuvor beschriebenen positiven Trends als Effekte einer Verbesserung der Bildungsqualität deuten lassen. So würden aktuelle Reformanstrengungen bestätigt, wenn sich die Verringerung der Leistungsnachteile von Jugendlichen mit Migrationshintergrund auf die deutlich verstärkten Bemühungen zurückführen ließen, die seit PISA 2000 im Bereich der Sprachförderung unternommen worden sind (vgl. zusammenfassend BaMF, 2007; Redder et al., 2010). Tatsächlich interpretiert sogar die OECD in ihrem internationalen Bericht die Veränderungen für Deutschland in diesem Sinne: „In response to the large inequities that the PISA 2000 assessment had revealed, German schools and states had invested heavily in disadvantaged students, including those from immigrant backgrounds“ (OECD, 2010, S. 107).

In der Tat wurden in der Folge des „PISA-Schocks“ vielerlei Initiativen vor allem in der Sprachförderung und Unterstützung von Kindern und Jugendlichen aus zugewanderten Familien auf den Weg gebracht, deren Effekte jedoch kaum empirisch anhand von belastbaren Evaluationsdesigns untersucht worden sind (vgl. Paetsch, Wolf & Stanat, 2010). Zudem besuchten die im Jahr 2009 getesteten Jugendlichen zum Zeitpunkt des „PISA-Schocks“ (Ende 2001) im Regelfall schon die zweite Jahrgangsstufe der Grundschule, können also von den Maßnahmen, die vor allem im Elementar- und Primarbereich ansetzen, nur sehr eingeschränkt erreicht worden sein. Hinsichtlich der Schul- und Unterrichtsqualität im Sekundarbereich zeigen die PISA-Befragungen, dass sich einige Indikatoren, wie etwa Schülerurteile zur Schüler-Lehrer-Beziehung und zur Disziplin im Unterricht, zwischen 2000 und 2009 verbessert haben (Klieme, Jude et al., 2010, S. 287). Es ist jedoch nicht möglich, Effekte dieser Veränderungen in der wahrgenommenen Prozessqualität zu quantifizieren und zu prüfen. Da also Ursachen im Bereich der pädagogischen Angebote und Maßnahmen nicht oder nur sehr schwer identifiziert werden können, ist die wissenschaftliche Suche nach Erklärungsansätzen für systemische Veränderungen auf theoretische Argumente und Abschätzungen angewiesen.

1.3 Schätzung möglicher Ursachen von Veränderungen zwischen PISA 2000 und PISA 2009

Ein Weg zur Abschätzung von Effekten pädagogischer Angebote und Maßnahmen ist die statistische Kontrolle von möglichen alternativen Einflussfaktoren. Diesen Weg beschreitet die OECD in ihrem internationalen Bericht über Trends in der Lesekompetenz zwischen 2000 und 2009 (OECD, 2010, S. 49ff.). Neben den beobachteten Differenzen zwischen beiden Messzeitpunkten werden auch adjustierte Werte berichtet, bei denen Veränderungen in der soziodemografischen Zusammensetzung der Schülerschaft herausgerechnet sind. Statistisch kontrolliert werden hier Alter, Geschlecht, sozioökonomischer Status, Migrationshintergrund und der Sprachgebrauch in der Fami-

lie.¹ Durch diese statistische Kontrolle erhöht sich für Deutschland der zwischen 2000 und 2009 zu verzeichnende Zuwachs in der mittleren Lesekompetenz von 13 Punkten auf der PISA-Skala (beobachteter Wert) auf 17.6 Punkte (adjustierter Wert; vgl. OECD, 2010, S. 153), und zur Erklärung heißt es: „In these countries, student performance would have increased more rapidly than the performance increase that was actually observed if the changes in the demographic and socio-economic composition of the student population had been accounted for. This is explained by the fact that within these countries, in 2009 the student population had a more disadvantaged background than in 2000“ (OECD, 2010, S. 50). Für Deutschland bedeutet dies: Wenn der Anteil der Schülerinnen und Schüler mit Migrationshintergrund 2000 gleich hoch gewesen wäre wie 2009, wäre der Zuwachs in der Lesekompetenz rechnerisch knapp fünf Punkte höher ausgefallen.² Die Problematik dieser Modellrechnung besteht jedoch darin, dass „Migrationshintergrund“ gerade in Deutschland im Jahr 2009 etwas anderes bedeutet als im Jahr 2000. Anstatt global den Anteil der Schülerinnen und Schüler mit „irgendeinem“ Migrationshintergrund zu kontrollieren, gilt es, verschiedene Migrantengruppen in einem differenzierten Modell zu unterscheiden. Dies soll in der vorliegenden Arbeit geleistet werden.

Im Folgenden soll ein Ansatz vorgestellt werden, der aufgrund theoretisch plausibler Annahmen abschätzt, welche Effekte auf das mittlere Kompetenzniveau mit bestimmten Faktoren verbunden sein könnten, die nicht unmittelbar mit der Schul- und Unterrichtsqualität oder mit pädagogischen Maßnahmen zusammenhängen. Solche Alternativerklärungen können beispielsweise die Bildungslaufbahn von Kindern und Jugendlichen betreffen oder sich auf Faktoren beziehen, die außerhalb des Bildungswesens anzusiedeln sind, wie etwa demografische Merkmale. Die Effekte dieser Faktoren sollen im Folgenden anhand von Modellrechnungen geschätzt werden. Mit dieser Art des Vorgehens arbeitet die Bildungswissenschaft – in Übereinstimmung mit dem Falsifikationskonzept der Wissenschaftstheorie – gezielt „gegen sich“: Es wird ausgelotet, wie weit mögliche Alternativerklärungen reichen könnten, d.h. welcher Teil des beobachteten Trends in der Lesekompetenz rechnerisch durch solche außerhalb der Schul- und Unterrichtsqualität angesiedelten Faktoren erklärbar ist.

Wichtige Aspekte der Bildungslaufbahn sind die Bildungsgänge, die von Kindern und Jugendlichen besucht werden, und in welcher Zeit sie diese durchlaufen. So wäre es denkbar, dass die Steigerung der Lesekompetenz, die zwischen PISA 2000 und PISA 2009 auftrat, auf Veränderungen in diesen Merkmalen zurückgeführt werden könnte

1 “The adjusted reading performance results (...) use the 2009 PISA sample as a reference. Thus, the results from previous assessments were adjusted to be comparable to the 2009 results. This was achieved by centring background characteristics on the 2009 average values for each country and then carrying out a regression with centred background characteristics to obtain adjusted trends. In other words, results for 2000, 2003 and 2006 were adjusted to match the 2009 data” (OECD, 2010, S. 139).

2 Die Korrektur des Leistungszuwachses schrumpft auf einen halben Punkt, wenn man nur für Alter und Geschlecht korrigiert (OECD, 2010, S. 144). Dies weist darauf hin, dass Veränderungen im Migrationshintergrund entscheidender sind.

(Klieme, Jude et al., 2010). Konkret geht es um die Verteilung auf Schulformen und Jahrgänge. Die Gymnasialbeteiligung erhöhte sich von 28.3 Prozent in PISA 2000 auf 33.5 Prozent in PISA 2009, während der Anteil der Jugendlichen im Hauptschul-Bildungsgang von 21.8 Prozent auf 19.1 Prozent abfiel. Die Anteile für die anderen Bildungsgänge (Realschule, Integrierte Gesamtschule, Berufsschule und Förderschulen) sind hingegen stabil geblieben (Ehmke & Jude, 2010). Aus nationalen Studien ist bekannt, dass die Lernentwicklung von Schülerinnen und Schülern in Gymnasien im Durchschnitt günstiger verläuft als in anderen Schulformen (z.B. Becker, Lüdtke, Trautwein & Baumert, 2006; Baumert, Stanat & Watermann, 2006). Demnach könnte sich eine erhöhte Gymnasialbeteiligung positiv auf die durchschnittliche Lesekompetenz von Schülerinnen und Schülern ausgewirkt haben.

In eine ähnliche Richtung geht die Frage, wie sich die Zielpopulation von PISA, also die Fünfzehnjährigen, auf Jahrgangsstufen verteilen. Wenn sich – wie nach dem „PISA-Schock“ vielfach propagiert – die Zahl der vorzeitigen Einschulungen erhöht, die Zahl der Zurückstellungen verringert und die Zahl der Klassenwiederholungen ebenfalls verringert hätte, sollte 2009 ein größerer Anteil der 15-jährigen Schülerinnen und Schüler die Jahrgangsstufe 10 erreicht haben als im Jahr 2000. Tatsächlich weisen die Daten der amtlichen Statistik in diese Richtung (Konsortium Bildungsberichterstattung, 2006, S. 235ff.). So haben seit 2003 acht Länder in Deutschland den Zeitpunkt der Einschulung vorverlegt (Autorengruppe Bildungsberichterstattung, 2012, S. 63). Auch anhand der PISA-Daten haben Klieme, Jude et al. (2010) ermittelt: „Im Durchschnitt sind die Jugendlichen im Jahr 2009 bei gleichem Alter etwa einen Monat in der Schulkarriere weiter als im Jahr 2000“ (S. 285). Sie interpretieren dies als Zuwachs im Umfang (Quantität) von Lernerfahrungen und vermuten, dass dieser Faktor zu den positiven Veränderungen auf Systemebene beigetragen haben könnte.

Eine weitere mögliche Ursache für die Veränderungen könnte in der Zusammensetzung der Population von Schülerinnen und Schülern mit Migrationshintergrund bestehen, die sich zwischen 2000 und 2009 teilweise deutlich verschoben hat (Stanat et al., 2010). So stieg der Anteil von Jugendlichen, die zur zweiten Generation (beide Elternteile im Ausland geboren, Kind in Deutschland geboren) zählen, über die Zeit statistisch signifikant von 5.1 Prozent auf 11.7 Prozent an. Gleichzeitig verringerte sich der Anteil von Jugendlichen der ersten Generation von 10.1 Prozent auf 5.8 Prozent. Die durchschnittliche Verweildauer der Kinder mit Migrationshintergrund in Deutschland hat sich also deutlich erhöht (Stanat et al., 2010, S. 215). Da eine längere Verweildauer von Kindern mit Migrationshintergrund mit verbesserten Sprachkenntnissen einhergehen dürfte, die eine wichtige Voraussetzung für Schulerfolg darstellen, kann angenommen werden, dass der Anteil der leistungsstärkeren Jugendlichen mit Migrationshintergrund über die Zeit zugenommen hat. Auch der relative Anteil der verschiedenen Herkunftsgruppen (z.B. Jugendliche, deren Familien aus der Türkei, aus Polen oder dem Gebiet der ehemaligen Sowjetunion zugewandert sind) hat sich über die Zeit verschoben, ebenso wie der Bildungshintergrund der Eltern von Jugendlichen mit Migrationshintergrund und der Anteil der Schülerinnen und Schüler, die in der Familie Deutsch sprechen. Es ist zu vermuten, dass ein Teil des Anstiegs der Lesekompetenz bei Jugend-

lichen mit Migrationshintergrund zwischen 2000 und 2009 durch diese Veränderungen in der Zusammensetzung der Gruppen erklärt werden kann.

Ähnlich argumentieren Bildungsökonominnen aus der Schweiz in einer Auswertung der Schweizer PISA-Daten (Cattaneo & Wolter, 2012). In dieser Arbeit wurden die Daten von Jugendlichen der ersten Generation verglichen mit Daten von Angehörigen der zweiten Einwanderergeneration. Zwischen 2000 und 2009 kam es in der Schweiz zu einem Anstieg der Lesekompetenz bei der zweiten Generation um 14 Punkte, bei der ersten Generation aber um 43 Punkte – ein Unterschied, den die Autoren als Effekt einer veränderten Einwanderungspolitik der Schweiz deuten. Da auch für Deutschland in der ersten Generation etwas stärkere Trendgewinne zu verzeichnen sind als in der zweiten Generation (vgl. Abschnitt 1.1), könnte dieses Argument in ähnlicher Weise für Deutschland gelten; es bedarf jedoch eigenständiger Abschätzungen der Größenordnung im nationalen Kontext.

Cattaneo und Wolter (2012) vermuten aufgrund von Modellrechnungen mit Hilfe der in der Ökonomie entwickelten Blinder-Oaxaca-Analyse (zur Methode siehe unten Abschnitt 3.3) für die Schweiz, dass etwa 70 Prozent des Leistungsanstiegs in der ersten Generation durch die Veränderung des sozioökonomischen Status der Einwanderer erklärt werden können (48 Prozent auf individueller Ebene, weitere 20 Prozent als Kompositionseffekt), die ihrerseits ein unmittelbares Ergebnis der schweizerischen Migrationspolitik sei, welche mittlerweile gezielt die Einwanderung hoch qualifizierter Personen bevorzuge.

2. Fragestellungen

Ausgehend von den beschriebenen Annahmen gehen wir im Folgenden der Frage nach, inwieweit sich der Anstieg in der Lesekompetenz zwischen PISA 2000 und PISA 2009 durch strukturelle Veränderungen im Bildungssystem erklären lässt. Als strukturelle Veränderungen bezeichnen wir dabei Verschiebungen in der Bildungslaufbahn der Fünfzehnjährigen hinsichtlich der Verteilungen auf Bildungsgänge und auf Jahrgangsstufen einerseits sowie Veränderungen in der Zusammensetzung der Population der Jugendlichen mit Migrationshintergrund andererseits.

Dabei erwarten wir, dass ein Teil des Leistungszuwachses durch den in PISA 2009 höheren Anteil von Schülerinnen und Schülern an Gymnasien, den verringerten Anteil von Jugendlichen an Hauptschulen und den höheren Anteil an Jugendlichen, die mindestens Jahrgangsstufe 10 erreicht haben, erklärt werden kann. Zudem sollte die über die Zeit veränderte Zusammensetzung der Migrantengruppe hinsichtlich des Generationsstatus, der Herkunftsländer, des Bildungshintergrunds und des Anteils von Familien mit deutscher Familiensprache zum mittleren Anstieg in der Lesekompetenz beigetragen haben. Die Höhe dieser erwarteten Effekte werden wir – dem Vorgehen von Cattaneo und Wolter (2012) für die Schweiz folgend – mit einer Blinder-Oaxaca-Analyse abschätzen.

3. Methodisches Vorgehen

3.1 Stichprobe

Die folgenden Analysen basieren auf den deutschen Teildatensätzen der internationalen Schulleistungsstudien PISA 2000 und PISA 2009. Die Stichprobe für PISA 2000 umfasst $N = 5012$ Schülerinnen und Schüler; für PISA 2009 liegen $N = 4979$ Fälle vor. Vertiefende Informationen zum Erhebungsdesign und zur Stichprobenziehung in PISA 2000 wurden im nationalen Berichtsband publiziert (Baumert, Stanat & Demmrich, 2001). Die entsprechenden Angaben für PISA 2009 finden sich bei Jude und Klieme (2010).

3.2 Erklärungsvariablen

Als Merkmale der Bildungslaufbahn betrachten wir die Verteilung der Jugendlichen auf die Bildungsgänge und unterscheiden zwischen Hauptschule, Realschule, Integrierte Gesamtschule, Gymnasium sowie andere Bildungsgänge (wie Berufsschulen und Förderschulen). Zudem wird die Jahrgangsstufe berücksichtigt, die die Fünfzehnjährigen zum Testzeitpunkt besucht haben. Es werden drei Kategorien differenziert: (1) Jahrgangsstufe 8 und darunter, (2) Jahrgangsstufe 9 sowie (3) Jahrgangsstufe 10 und höher.

Hinsichtlich des Migrationshintergrunds der Jugendlichen unterscheiden wir analog zur nationalen Berichterstattung in PISA 2009 die folgenden Gruppen (Stanat et al., 2010): (1) „ohne Migrationshintergrund“: kein Elternteil im Ausland geboren, (2) „ein Elternteil im Ausland geboren“: ein Elternteil im Ausland, ein Elternteil in Deutschland geboren, (3) „zweite Generation“: beide Elternteile im Ausland geboren, Jugendlicher in Deutschland geboren und (4) „erste Generation“: beide Elternteile und Jugendlicher im Ausland geboren.

Darüber hinaus werden in den Analysen drei Herkunftsgruppen unterschieden, die wie folgt definiert sind: (1) Herkunftsland ehemalige UdSSR: mindestens ein Elternteil in Russland, Kasachstan oder einer anderen ehemaligen Sowjetrepublik geboren, (2) Herkunftsland Türkei: mindestens ein Elternteil in der Türkei geboren und (3) Herkunftsland Polen: mindestens ein Elternteil in Polen geboren.

Die Kodierung des höchsten Bildungsabschlusses der Eltern innerhalb der Migrationsgruppen erfolgt anhand der *International Standard Classification of Education* (ISCED). Wir unterscheiden dabei vier Niveaus: Bildungsabschluss im Sekundarbereich I oder darunter (ISCED 1, 2), Abschluss im Sekundarbereich II oder im post-sekundaren Bereich (ISCED 3, 4), Abschluss im nichtuniversitären Tertiärbereich (ISCED 5B) und Abschluss im universitären Tertiärbereich (ISCED 5A, 6).

Die Familiensprache der Schülerinnen und Schüler wurde mit einem Fragebogenitem erfasst („Welche Sprache spricht ihr normalerweise zu Hause?“) und geht als dichotome Variable in die Analysen ein (Familiensprache Deutsch vs. Familiensprache nicht Deutsch).

Fehlende Werte für den Migrationshintergrund wurden wie im nationalen Bericht zu PISA 2009 behandelt (vgl. Stanat et al., 2010). Zusätzlich zu den oben beschriebenen Gruppen wurde eine weitere Kategorie gebildet, die als „nicht zuzuordnen“ bezeichnet wird. Dabei handelt es sich um Jugendliche, für die keine vollständigen Informationen zum eigenen Geburtsland oder zum Geburtsland der Eltern vorliegen. Diese Gruppe wurde als eigene Kategorie mit in die Analysen und in die Ergebnistabellen aufgenommen.

3.3 *Blinder-Oaxaca-Analyse*

Als statistische Methode verwenden wir die Blinder-Oaxaca-Decomposition-Technik, die bislang überwiegend in den Wirtschaftswissenschaften angewendet wird. Diese Analyse wurde ursprünglich von Blinder (1973) und Oaxaca (1973) entwickelt, um die Lohnunterschiede zwischen Frauen und Männern zu erklären. Die Mittelwertsdifferenz einer Variablen (z.B. der Lohn) zwischen zwei Gruppen (z.B. Frauen und Männer) wird dabei in zwei Teile zerlegt. Der erste Teil erklärt die Mittelwertsdifferenz anhand von Merkmalsunterschieden zwischen den beiden Gruppen; der zweite Teil beschreibt die Mittelwertsdifferenz anhand von Unterschieden in Regressionskoeffizienten und Achsenabschnitten. Während also der erste Teil abbildet, welchen Einfluss die unterschiedliche Zusammensetzung der zu vergleichenden Gruppen hinsichtlich weiterer Hintergrundmerkmale (z.B. Erwerbstätigkeitsstatus) hat, wird mit dem zweiten erfasst, inwieweit mit diesen weiteren Hintergrundmerkmalen für die untersuchten Gruppen (ursprünglich: Frauen und Männer) unterschiedliche Effekte auf die Zielvariable (ursprünglich: Lohnhöhe) verbunden sind.

In unserer Analyse untersuchen wir die Veränderung in der Lesekompetenz der Fünfzehnjährigen von PISA 2000 nach PISA 2009. Analog zu Blinder (1973) und Oaxaca (1973) bilden die beiden Zeitpunkte 2000 und 2009 die Gruppen, zwischen denen die Mittelwertsunterschiede in den Testleistungen, T_{00} und T_{09} , aufzuklären sind. In einem ersten Schritt sind dazu zwei multiple Regressionen zu berechnen.

- (1) $T_{09i} = b_{09} X_{09i} + \varepsilon_{09i}$
- (2) $T_{00i} = b_{00} X_{00i} + \varepsilon_{00i}$

Die Testleistungen T_{00i} in PISA 2000 und T_{09i} in PISA 2009 werden für jede Person i durch einen Vektor X von potentiellen Prädiktorvariablen vorhergesagt. Zu den Prädiktoren können beispielsweise der Besuch verschiedener Bildungsgänge und Jahrgangsstufen sowie Merkmale des Migrations- und Bildungshintergrunds zählen. Die Vektoren b_{00} und b_{09} repräsentieren die jeweiligen Achsenabschnitte und Regressionskoeffizienten, die Variablen ε_{00i} und ε_{09i} sind die Fehlerterme der Regressionsgleichungen.

Die Mittelwertsdifferenz zwischen beiden Zeitpunkten lässt sich dann ausdrücken als:

$$(3) \quad \bar{T}_{09} - \bar{T}_{00} = \hat{\beta}_{09} \bar{X}_{09} - \hat{\beta}_{00} \bar{X}_{00},$$

wobei \bar{T}_{09} und \bar{T}_{00} die Mittelwerte in der Lesekompetenz und $\hat{\beta}_{09}$ und $\hat{\beta}_{00}$ die geschätzten Regressionskoeffizienten aus den Gleichungen (1) und (2) sind (vgl. Zhang & Lee, 2011). Nach Blinder und Oaxaca kann diese Differenz nun wie folgt umgeformt werden:

$$\begin{aligned} (4) \quad \bar{T}_{09} - \bar{T}_{00} &= \hat{\beta}_{09} \bar{X}_{09} - \hat{\beta}_{00} \bar{X}_{00} \\ &= \hat{\beta}_{09} \bar{X}_{09} - \hat{\beta}_{00} \bar{X}_{00} + \hat{\beta}_{09} \bar{X}_{00} - \hat{\beta}_{09} \bar{X}_{00} \\ &= \hat{\beta}_{09} (\bar{X}_{09} - \bar{X}_{00}) + (\hat{\beta}_{09} - \hat{\beta}_{00}) \bar{X}_{00} \end{aligned}$$

Der Zuwachs in der Lesekompetenz von PISA 2000 nach PISA 2009 ergibt sich somit aus einer Summe von zwei Termen. Der linke Teil des Summenterms der Gleichung (4), $\hat{\beta}_{09} (\bar{X}_{09} - \bar{X}_{00})$, gewichtet die Differenzen in den Prädiktorvariablen mit den Regressionskoeffizienten aus PISA 2009. Der Anstieg in der Lesekompetenz über die Zeit wird somit durch die Veränderung in den Prädiktorvariablen von 2000 nach 2009 erklärt. Inhaltlich beschreibt dieser Teil, wie viel Zuwachs an Lesekompetenz sich durch die Veränderung der Merkmalsausprägungen zwischen 2000 und 2009 ergeben würde. Dieser Term wird auch als aufgeklärter Anteil bezeichnet.

Der rechte Teil des Summenterms beschreibt den Beitrag der Veränderungen in den Regressionskoeffizienten über die Zeit für den Mittelwertsunterschied in der Lesekompetenz. Inhaltlich wird damit der Effekt erfasst, der mit der über die Zeit veränderten Vorhersagekraft der Prädiktoren verbunden ist. Dieser Anteil schließt alle Einflüsse mit ein, die nicht durch die Veränderung in den beobachteten Populationsmerkmalen abgedeckt sind. Diese Komponente wird daher auch als unerklärter Anteil oder „discrimination effect“ bezeichnet (Zhang & Lee, 2011).

4. Ergebnisse

Die Ergebnisse der Blinder-Oaxaca-Analyse sind in der Tabelle 1 zusammengefasst. Den theoretischen Überlegungen folgend haben wir fünf Merkmalsbereiche in die Analyse einbezogen: (1) Verteilung der Schülerinnen und Schüler auf Bildungsgänge, (2) Anteile der Fünfzehnjährigen auf verschiedenen Jahrgangsstufen, (3) Migrationsgruppen nach Generationsstatus und Herkunftsländern, (4) Migrationsgruppen nach Generationsstatus und Bildungsabschluss der Eltern und (5) Migrationsgruppen nach Generationsstatus und Familiensprache. Für alle Einzelprädiktoren sind in der Tabelle 1 die absoluten Prozentanteile (Spalten 2 und 3) sowie die Veränderungen in den jeweiligen Anteilen zwischen PISA 2000 und 2009 angegeben (vgl. Spalte 4). Die Spalten 5-7 enthalten die unstandardisierten Regressionskoeffizienten für die Vorhersage der Lesekompetenz zu beiden Zeitpunkten sowie deren Differenz. Letztere beschreibt, inwie-

weit die Vorhersagekraft der einzelnen Prädiktoren für die Lesekompetenz über die Zeit gestiegen ist oder abgenommen hat. Die Beiträge der einzelnen Merkmale stehen in den Spalten 8 und 9. Die Summe aller Einzelbeiträge ergibt den Gesamtzuwachs in der Lesekompetenz von 13 Punkten. Im Folgenden werden die Ergebnisse für die fünf Merkmalsbereiche nacheinander vorgestellt.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Variable	\hat{X}_{00}	\hat{X}_{09}	$\hat{X}_{09} - \hat{X}_{00}$	$\hat{\beta}_{00}$	$\hat{\beta}_{09}$	$(\hat{\beta}_{09} - \hat{\beta}_{00})$	$\hat{\beta}_{09}(\bar{X}_{09} - \bar{X}_{00})$	$(\hat{\beta}_{09} - \hat{\beta}_{00})\bar{X}_{00}$	
Bildungsgänge (Referenz: Realschule)									
Hauptschule	21.8%	19.1%	-2.7%	-74.7	-68.3	6.4	1.8	1.4	
Integrierte Gesamtschule	8.8%	8.7%	-0.1%	-35.9	-18.5	17.4	0.0	1.5	
Gymnasium	28.3%	33.5%	5.2%	76.7	65.9	-10.8	3.4	-3.1	
Andere Bildungsgänge	8.7%	7.7%	-1.0%	-87.0	-108.5	-21.5	1.1	-1.9	
Jahrgangsstufe (Referenz: Jahrgangsstufe 9)									
Jahrgangsstufe 8 oder darunter	15.7%	11.7%	-4.0%	-40.5	-34.5	6.0	1.4	0.9	
Jahrgangsstufe 10 oder darüber	24.7%	35.7%	11.0%	41.4	37.9	-3.4	4.2	-0.8	
Migrationshintergrund (Generationenstatus und Herkunftsland)									
(Referenz: Kinder ohne Migrationshintergrund)									
Ein Elternteil im Ausland geboren, Polen	0.6%	0.8%	0.1%	-39.2	-77.3	-38.2	-0.1	-0.2	
Ein Elternteil im Ausland geboren, Türkei	0.3%	0.9%	0.6%	-66.0	-106.1	-40.1	-0.6	-0.1	
Ein Elternteil im Ausland geboren, UdSSR	0.3%	0.3%	0.1%	-57.6	-81.0	-23.4	0.0	-0.1	
Ein Elternteil im Ausland geboren, anderes Land	4.9%	5.3%	0.4%	-33.8	-81.1	-47.3	-0.3	-2.3	
1. Generation, Polen	1.5%	0.3%	-1.2%	-28.1	-28.6	-0.5	0.3	0.0	
1. Generation, Türkei	0.8%	0.5%	-0.3%	-54.6	-38.6	16.1	0.1	0.1	
1. Generation, UdSSR	4.3%	2.6%	-1.8%	-29.2	-25.1	4.1	0.4	0.2	
1. Generation, Anderes Land	3.0%	1.8%	-1.2%	-43.8	-32.0	11.8	0.4	0.3	
2. Generation, Polen	0.2%	1.4%	1.1%	-46.2	-36.9	9.3	-0.4	0.0	
2. Generation, Türkei	1.9%	3.7%	1.8%	-49.8	-66.1	-16.3	-1.2	-0.3	
2. Generation, UdSSR	0.2%	1.9%	1.6%	-25.9	-31.0	-5.2	-0.5	0.0	
2. Generation, anderes Land	2.5%	3.5%	1.0%	-21.8	-50.4	-28.6	-0.5	-0.7	
Migrationshintergrund (Generationenstatus und Bildungsabschluss Eltern)									
(Referenz: Kinder ohne Migrationshintergrund)									
Ein Elternteil im Ausland geboren, ISCED 1,2	0.7%	0.9%	0.2%	6.7	32.3	25.7	0.1	0.2	
Ein Elternteil im Ausland geboren, ISCED 3,4	2.2%	2.5%	0.3%	32.8	49.8	17.0	0.1	0.4	
Ein Elternteil im Ausland geboren, ISCED 5A,6	2.0%	2.4%	0.3%	26.6	51.9	25.2	0.2	0.5	
Ein Elternteil im Ausland geboren, ISCED 5B	0.8%	1.1%	0.2%	18.7	43.8	25.1	0.1	0.2	
1. Generation, ISCED 1,2	2.0%	0.9%	-1.1%	-25.0	-45.1	-20.1	0.5	-0.4	
1. Generation, ISCED 3,4	1.5%	1.2%	-0.3%	-4.3	2.4	6.8	0.0	0.1	
1. Generation, ISCED 5A,6	2.3%	1.4%	-0.9%	-11.3	-4.7	6.7	0.0	0.2	
1. Generation, ISCED 5B	0.9%	0.6%	-0.2%	23.6	-4.7	-28.4	0.0	-0.2	
2. Generation, ISCED 1,2	1.4%	2.4%	1.0%	-29.2	4.2	33.4	0.0	0.5	
2. Generation, ISCED 3,4	1.0%	3.3%	2.3%	-9.8	9.1	18.9	0.2	0.2	
2. Generation, ISCED 5A,6	1.0%	1.7%	0.6%	-19.7	6.7	26.4	0.0	0.3	
2. Generation, ISCED 5B	0.3%	1.1%	0.8%	-27.1	16.0	43.1	0.1	0.1	
Migrationshintergrund (Generationenstatus und Familiensprache)									
(Referenz: Kinder ohne Migrationshintergrund)									
Ein Elternteil im Ausland geboren, Familiensprache deutsch	4.4%	5.8%	1.5%	9.0	39.6	30.6	0.6	1.3	
Ein Elternteil im Ausland geboren, Familiensprache nicht deutsch	0.3%	0.7%	0.4%	-0.2	25.8	26.0	0.1	0.1	
1. Generation, Familiensprache deutsch	2.0%	1.1%	-0.9%	30.7	21.0	-9.7	-0.2	-0.2	
1. Generation, Familiensprache nicht deutsch	4.3%	3.1%	-1.2%	-14.0	12.6	26.6	-0.2	1.2	
2. Generation, Familiensprache deutsch	1.1%	4.2%	3.1%	35.2	28.9	-6.4	0.9	-0.1	
2. Generation, Familiensprache nicht deutsch	1.6%	4.2%	2.7%	9.1	11.7	2.6	0.3	0.0	
Migrationshintergrund (fehlende Werte)	5.2%	11.0%	5.8%	-131.7	-26.1	105.6	-1.5	5.5	
Lesekompetenz (Spalten 2-4) / Intercept (Spalten 5-7)	484	497	13.3	498.7	496.2	-2.5	-2.5		
							10.9	2.4	13.3
							82.1%	17.9%	100%
							Explained	Unexplained	Gap

zu bestimmen, werden diese Differenzen in den Prozentanteilen mit den Regressionskoeffizienten aus PISA 2009 multipliziert (Spalte 8). Inhaltlich drückt dies aus, um wie viele Punkte sich der nationale Durchschnittswert zwischen PISA 2000 und PISA 2009 rechnerisch erhöhen (oder verringern) würde, wenn die prozentuale Zusammensetzung nach Bildungsgängen in PISA 2000 wie in PISA 2009 ausgefallen wäre und die Gruppen jeweils dieselbe mittlere Lesekompetenz erzielt hätten wie in PISA 2009. Addiert man die Kennwerte in Spalte 8 für die einzelnen Bildungsgänge zusammen, so zeigt sich, dass diese Verschiebungen in den Anteilen der Bildungsgänge einem Anstieg in der Lesekompetenz von 6.3 Punkten entsprechen würden (Hauptschule: +1.8 Punkte, IGS: +0 Punkte, Gymnasium: +3.4 Punkte, andere Bildungsgänge: +1.1 Punkte).

Anteile von Fünfzehnjährigen auf unterschiedlichen Jahrgangsstufen

Ein weiterer wichtiger Aspekt der Bildungslaufbahn ist deren zeitlicher Verlauf. Im Vergleich zwischen 2000 und 2009 zeigt PISA, dass der Anteil derjenigen Schülerinnen und Schüler, die noch Jahrgangsstufe 8 oder eine niedrigere Jahrgangsstufe besuchen, um 4 Prozentpunkte gesunken, der Anteil in Jahrgangsstufe 10 oder höher hingegen um 11 Prozentpunkte gestiegen ist (vgl. Spalte 2 und 3). Geht man also davon aus, dass 15 Prozent der Fünfzehnjährigen 2009 eine Jahrgangsstufe weiter gekommen waren als 2000, heißt dies: Im Durchschnitt sind die Jugendlichen im Jahr 2009 bei gleichem Alter etwa zwei Monate in der Schulkarriere weiter als im Jahr 2000.³ Geht man – wiederum mit Baumert et al. (2006) sowie OECD (2004) – davon aus, dass der Lerngewinn im Lesen im Laufe eines Schuljahres etwa 30 Punkten beträgt, sollte dieser Faktor insgesamt mit etwa 4 bis 5 Punkten beim nationalen Durchschnittswert der Lesekompetenz zu Buche schlagen.

Eine genauere Abschätzung dieses Effekts zeigen die Kennwerte in der Spalte 8 für die Jahrgangsstufe 8 und 10. Demnach könnten 5.6 Punkte des Zuwachses in der Lesekompetenz durch die veränderten Prozentanteile von Schülerinnen und Schülern auf

3 Zu einem gewissen Teil ist die Änderung in der Verteilung auf Jahrgangsstufen ihrerseits zurückführbar auf eine technische Anpassung der Stichprobendefinition, die im Jahr 2003 erfolgte. Zur PISA-Zielpopulation von 2000 gehörten alle zwischen Dezember 1984 und November 1985 Geborenen. Zur PISA-Zielpopulation von 2003 gehörten alle zwischen Januar 1988 und Dezember 1988 Geborenen. Parallel dazu wurde das Testzeitfenster entsprechend um einen Monat verschoben. Wie international gefordert, hat sich also an der Altersverteilung der Getesteten nichts geändert. Grund für diese Änderung waren technische Vorteile: Das veränderte Testfenster passte besser zu den Schulferien, und die Ziehung der Stichprobe war einfacher, weil man nur das Geburtsjahr, nicht aber den Geburtsmonat kennen musste. Diese Definition der Zielpopulation wurde seit 2003 in allen PISA-Erhebungen verwendet. Ein Nebeneffekt war allerdings, dass wegen der Stichtagsregelung bei der Einschulung sich automatisch ein gewisser Teil der Schülerschaft in einer höheren Jahrgangsstufe befinden würde, und zwar theoretisch $\frac{1}{12}$ der Stichprobe. Somit geht die Veränderung der Verteilung auf Jahrgangsstufen schätzungsweise hälftig auf tatsächlich veränderte Bildungswege (frühere Einschulung, weniger Klassenwiederholungen) zurück und hälftig auf die gegenüber 2000 abgewandelte Definition der Zielpopulation.

den Jahrgangsstufen zurückzuführen sein (Jahrgangsstufe 8: +1.4 Punkte, Jahrgangsstufe 10: +4.2 Punkte).

Veränderungen in der Zusammensetzung der Migrantengruppe über die Zeit

Als weiteren Merkmalsbereich nehmen wir den Migrationshintergrund der Schülerinnen und Schüler in den Blick, da wir wissen, dass Jugendliche aus zugewanderten Familien maßgeblich zum Trendgewinn in Deutschland beigetragen haben (vgl. Abschnitt 1.1). Wir untersuchen im Folgenden, inwieweit der Anstieg in der Lesekompetenz bei diesen Jugendlichen durch eine veränderte Zusammensetzung der Migrantengruppe erklärt werden kann. Um diese Vermutung zu überprüfen, wird die Gruppe der Jugendlichen mit Migrationshintergrund differenziert in die Analyse einbezogen. Dabei werden gleichzeitig die Effekte einer veränderten Zusammensetzung der Gruppe hinsichtlich der Herkunftsländer und des Generationenstatus, des Bildungsabschlusses und der Familiensprache analysiert. Anders als Cattaneo und Wolter (2012) betrachten wir nicht den sozioökonomischen Status der zugewanderten Familien, da sich dieser in Deutschland zwischen 2000 und 2009 kaum verändert hat (vgl. Klieme, Jude et al., 2010, S. 283, oder Stanat et al., 2010, S. 219), sondern den Zeitpunkt ihrer Zuwanderung, ihr Herkunftsland sowie ihren Bildungshintergrund und die in der Familie gesprochene Sprache.

Die deskriptiven Kennwerte für Schülerinnen und Schüler mit Migrationshintergrund, differenziert nach Generationenstatus und Herkunftsland, geben die erwähnten Verschiebungen in den prozentualen Anteilen noch einmal wieder: in PISA 2009 ist der Anteil der Jugendlichen der zweiten Generation gestiegen, während der Anteil der Kinder der ersten Generation abgenommen hat. Diese Tendenz bestätigt sich für alle Gruppen aus verschiedenen Herkunftsländern (vgl. Spalte 4).

Die Kennwerte in der Spalte 8 beschreiben nun, um wie viele Punkte die mittlere Lesekompetenz in der Gesamtstichprobe ansteigen würde, wenn die prozentuale Zusammensetzung der zwölf Migrantengruppen in PISA 2000 wie in PISA 2009 ausgefallen wäre. Fasst man die Werte für die vier Migrationsgruppen der ersten Generation zusammen, so ergibt sich ein Anstieg von +1.2 Punkten in der Lesekompetenz. Für die vier Herkunftsgruppen der zweiten Generation ergibt sich hingegen eine Verringerung der mittleren Lesekompetenz um –2.6 Punkte. Auch für die Gruppe der Jugendlichen, bei denen ein Elternteil im Ausland geboren wurde, wäre durch die Merkmalsverschiebung eine Abnahme der Lesekompetenz um –1.0 Punkte zu erwarten.

Neben den Migrationsgruppen differenziert nach Herkunftsländern untersuchen wir, inwieweit die Veränderung der Lesekompetenz auf die veränderte Zusammensetzung der Jugendlichen mit Migrationshintergrund hinsichtlich des elterlichen Bildungsabschlusses zurückgeführt werden kann. Auch damit soll geprüft werden, inwieweit sich die Steigerung der Lesekompetenz über die Zeit durch Verschiebungen in der Struktur der Migrantengruppen erklären lässt. In Tabelle 1 sind die nach elterlichem Bildungsabschluss und Generationenstatus differenzierten absoluten Prozentanteile von Schülerinnen und Schülern mit Migrationshintergrund angegeben.

Welcher Anstieg in der Lesekompetenz würde sich nun für die Jugendlichen ergeben, wenn in PISA 2000 die prozentuale Zusammensetzung der nach Bildungsabschluss differenzierten Migrantengruppen wie in PISA 2009 ausgefallen wäre? Die Abschätzung dieses Effekts ist wiederum durch die Kennwerte in der Spalte 8 gegeben. Diese fallen insgesamt geringer aus als für die Differenzierung der Migrationsgruppen nach Herkunftsstaaten. Fasst man die Gruppen nach Generationenstatus zusammen, so ergibt sich für die veränderten Anteile in der ersten Generation ein Anstieg von +0.5 Punkten und für die zweite Generation von +0.3 Punkten. Demnach scheint die strukturelle Veränderung im elterlichen Bildungsabschluss bei Jugendlichen mit Migrationshintergrund nur in geringfügigem Maße zum Anstieg in der Lesekompetenz beigetragen zu haben.

Die letzte Gruppe von Variablen berücksichtigt die zwischen PISA 2000 und PISA 2009 veränderte Zusammensetzung der Migrantengruppen hinsichtlich der zu Hause gesprochenen Familiensprache. In der Tabelle 1 werden dazu die Migrantengruppen nach Generationenstatus unterteilt in Familien, in denen überwiegend Deutsch gesprochen wird, und in Familien, in denen überwiegend eine andere Herkunftssprache im Alltag genutzt wird. Fasst man die Ergebnisse für diese Variablen zusammen, so ergibt sich ein Anstieg von +1.5 Punkten in der Lesekompetenz. Verglichen mit den Ergebnissen der beiden vorherigen Merkmale der Schülerinnen und Schüler mit Migrationshintergrund kommt der Familiensprache damit eine etwas höhere Erklärungskraft zu.

Die vorletzte Zeile in der Tabelle 1 gibt jeweils den Anteil in den Stichproben an, für den sich der Migrationshintergrund aufgrund fehlender Werte nicht eindeutig bestimmen lässt. Dieser Anteil ist über die Zeit von 5.2 Prozent auf 11 Prozent bedeutsam angestiegen. Auch diese Verschiebung muss berücksichtigt werden, um die Effekte in den veränderten Populationsmerkmalen auf die Lesekompetenz über die Zeit adäquat abzubilden. Spalte 8 gibt wiederum an, wie sich die Lesekompetenz in der Gesamtgruppe aufgrund der Anteilsverschiebung für dieses Merkmal verändern würde. Demnach würde die Mittelwertsdifferenz in der Lesekompetenz zwischen PISA 2000 und PISA 2009 um 1.5 Punkte geringer ausfallen, wenn der Anteil der Schülerinnen und Schüler mit fehlenden Werten auf der Migrationsvariablen 2000 so groß gewesen wäre wie 2009 und wenn die mittlere Leistung dieser Gruppe der in 2009 beobachteten entspräche. Gleichzeitig wird aber auch deutlich, dass sich die Vorhersagekraft der fehlenden Angaben zum Migrationshintergrund für die Lesekompetenz gemessen an der Höhe der Regressionskoeffizienten zwischen den beiden Zeitpunkten deutlich unterscheidet (Spalte 5 und 6). Entsprechend hoch fällt auch der Beitrag aus (5.5 Punkte), der auf Differenzen in den Regressionskoeffizienten zurückgeführt werden kann (Spalte 9). Dies weist darauf hin, dass sich die Bedeutung dieser Variablen über die Zeit offensichtlich geändert hat: Jugendliche mit fehlenden Angaben zum Migrationshintergrund erreichten in PISA 2000 eine deutlich niedrigere Lesekompetenz als entsprechende Schülerinnen und Schüler in PISA 2009. Eine inhaltliche Interpretation dieses Terms ist daher nicht möglich. Um eine unverzerrte Schätzung der Gesamtmittelwerte zu gewährleisten, müssen diese Schülerinnen und Schüler aber in der Analyse mit berücksichtigt werden.

Ergebnis der Blinder-Oaxaca-Analyse

Der Anstieg in der Lesekompetenz zwischen PISA 2000 und PISA 2009 für die Schülerinnen und Schüler aus Deutschland beträgt 13.3 Punkte. Diese Punktedifferenz kann nach Oaxaca in zwei Teile zerlegt werden: in einen Anteil, der durch veränderte Merkmalsausprägungen erklärt wird (vgl. Spalte 8), und in einen Anteil, der auf veränderte Wirkzusammenhänge zurückzuführen ist (vgl. Spalte 9). Summiert man nun alle Werte der Spalten 8 auf, ergibt sich der Anteil an Punkten in der Lesekompetenz, der auf die veränderten Merkmalsbedingungen zurückgeführt werden kann. Nach unserer Analyse sind dies 10.9 Punkte, d.h. rund 82 Prozent des Anstiegs in der Lesekompetenz. Entsprechend verbleiben noch etwa 2.4 Punkte (ca. 18 Prozent), die auf andere Merkmale beziehungsweise veränderte Wirkzusammenhänge zurückgeführt werden können und die durch das Modell nicht erklärt werden.

5. Diskussion

Das Ziel dieses Beitrags bestand darin, abzuschätzen, inwieweit strukturelle Veränderungen in der Bildungsbeteiligung (Verteilung der Fünfzehnjährigen auf Bildungsgänge und Jahrgangsstufen) sowie veränderte soziodemografische Bedingungen (Zusammensetzung der Gruppe der Jugendlichen mit Migrationshintergrund) den Anstieg in der Lesekompetenz bei Jugendlichen in Deutschland verursacht haben könnten.

Methodisch haben wir uns an dem Vorgehen von Cattaneo & Wolter (2012) orientiert und das in der empirischen Bildungsforschung bislang noch wenig genutzte Verfahren der Blinder-Oaxaca-Analyse angewendet. Während diese Methode in den Wirtschaftswissenschaften weitgehend etabliert ist, gibt es in der empirischen Bildungsforschung bislang erst wenige Arbeiten, die diese Methode nutzen (z.B. Zhang & Lee, 2011; Ammermueller, 2007). Der Anstieg in der Lesekompetenz zwischen PISA 2000 und PISA 2009 wurde dabei zerlegt in einen Anteil, der durch Veränderungen in der Bildungsbeteiligung und der Bildungslaufbahn sowie in soziodemografischen Merkmalen der Schülerschaft erklärt wird, und in einen Anteil, der auf veränderte Wirkzusammenhänge zurückzuführen ist. Durch unsere Analysen konnten insgesamt 10.9 von 13.3 Punkten (rund 82 Prozent) der Verbesserung im Lesen auf veränderte Ausprägungen in den untersuchten Merkmalen zurückgeführt werden. Entsprechend verblieben noch 2.4 Punkte (ca. 18 Prozent) des Anstiegs, die sich nicht anhand von diesen Variablen erklären lassen und somit auf eine veränderte pädagogische Praxis zurückzuführen sein könnten.

Von den einzelnen strukturellen Merkmalen, die in der Analyse gleichzeitig berücksichtigt wurden, ließ sich erwartungskonform ein positiver Erklärungsbeitrag für die über die Zeit gestiegene Bildungsbeteiligung finden. Nach der Blinder-Oaxaca-Analyse konnte ein Teil des Anstiegs in der Lesekompetenz, nämlich etwa 6 von 13 Punkten, auf die erhöhte Gymnasialquote sowie den gesunkenen Anteil von Hauptschülern zurückgeführt werden.

Ebenfalls war anzunehmen, dass der gestiegene Anteil von Fünfzehnjährigen auf höheren Jahrgangsstufen (mehr Schülerinnen und Schüler in Jahrgangsstufe 10 und

weniger in Jahrgangsstufe 8) einen Teil des Anstiegs in der Lesekompetenz erklärt. Auch diese Vermutung konnte bestätigt werden. Auf dieses Merkmal konnten weitere 6 Punkte in der Veränderung der Lesekompetenz zurückgeführt werden.

Die strukturellen Verschiebungen in der Zusammensetzung der Jugendlichen mit Migrationshintergrund haben wir in der Analyse sehr differenziert abgebildet und die Veränderungen für drei Merkmale berücksichtigt: Herkunftsländer, Bildungsabschlüsse der Eltern und Familiensprache. Dabei war die Vermutung, dass eine veränderte Zusammensetzung in der Gruppe der Jugendlichen mit Migrationshintergrund den Anstieg in der Lesekompetenz über die Zeit zumindest teilweise erklären kann.

In unserer nach Generationenstatus und Herkunftsländern differenzierten Analyse zeigte sich, dass der gestiegene Anteil von Jugendlichen der zweiten Generation mit einer Verringerung der mittleren Lesekompetenz um 2.6 Punkte verbunden war. Hingegen hatte der verringerte Anteil von Jugendlichen der ersten Generation einen positiven Effekt auf den Zuwachs in der Lesekompetenz (+1.2 Punkte). Dieser auf den ersten Blick (vielleicht) kontraintuitive Befund lässt sich dadurch erklären, dass die Jugendlichen der ersten Generation im Mittel eine unterdurchschnittliche Lesekompetenz aufweisen. Wenn der Anteil dieser Subpopulation insgesamt abnimmt, erhöht sich dadurch der Mittelwert in der Gesamtpopulation. Bei den Jugendlichen mit Migrationshintergrund der zweiten Generation ist es ähnlich. Auch diese Subpopulation erreichte im Durchschnitt ein Niveau in der Lesekompetenz, das unterhalb des Gesamtdurchschnitts lag. Der prozentuale Anteil dieser Subpopulation ist über die Zeit angestiegen. In der Gesamtpopulation führt dies daher zu einer (schwachen) Verringerung der mittleren Lesekompetenz.

Veränderungen im elterlichen Bildungsabschluss bei Jugendlichen mit Migrationshintergrund haben nach unserer Analyse nicht bedeutsam zum Anstieg in der Lesekompetenz beigetragen. Hingegen kam dem über die Zeit gestiegenen Anteil von Jugendlichen, die zu Hause überwiegend Deutsch sprechen, eine größere Bedeutung zu. Auf den gestiegenen Anteil von Familien mit Migrationshintergrund, in denen Deutsch als Familiensprache gesprochen wird, konnten 1.5 Punkte des Zuwachses in der Lesekompetenz zurückgeführt werden.

Insgesamt haben wir in unserer Analyse strukturelle Merkmalsbereiche identifiziert, die einen Teil des Zuwachses in der Lesekompetenz von Jugendlichen zwischen 2000 und 2009 erklären können. Die Ergebnisse lassen sich als Hinweis darauf werten, dass der Anstieg in der Lesekompetenz von Jugendlichen in Deutschland zu etwa 82 Prozent durch Veränderungen in der Bildungsbeteiligung, in der Verteilung auf Jahrgangsstufen und durch Veränderungen in der Verteilung von Hintergrundmerkmalen der Schülerinnen und Schüler mit Migrationshintergrund, vor allem im Sprachgebrauch in den Familien, verursacht war.

Entsprechend geringer fällt – folgt man der Logik dieser Modellrechnungen – der Anteil aus, der potenziell auf Reformen im Bildungssystem oder auf spezifische Förderung in Schule und Elternhaus zurückgeführt werden kann. Allerdings muss berücksichtigt werden, dass einige der hier als „exogen“ betrachteten Faktoren, insbesondere vermehrter Gymnasialbesuch sowie mehr Schülerinnen und Schüler im Alter von 15 Jahren

auf höheren Jahrgangsstufen, durchaus auch durch eine veränderte Praxis in Unterricht und Schule beeinflusst sein können. So könnte die über die Zeit gestiegene Gymnasialquote darauf beruhen, dass die an PISA 2009 teilnehmenden Schülerinnen und Schüler möglicherweise bereits am Ende ihrer Grundschulzeit höhere Kompetenzen besaßen, was sich dann später auch in den PISA-Tests niederschlug. Insofern bestätigt sich das im nationalen Bericht zu PISA 2009 gezogene Fazit: „Im Endeffekt muss offenbleiben, welche bildungspolitischen Maßnahmen und Programme zu Entwicklungen im vergangenen Jahrzehnt beigetragen haben und wie. Dass die Effekte in Deutschland vielfältig und eher unspezifisch sind, spricht dafür, dass letztlich die Kombination vieler unterschiedlicher Ansätze erforderlich war“ (Klieme, Jude et al., 2010, S. 291).

Bei der Interpretation der Ergebnisse der Blinder-Oaxaca-Analyse ist aus methodischer Sicht zu berücksichtigen, dass es sich dabei um ein Zerlegungsverfahren handelt, das auf einigen grundlegende Annahmen basiert. So gehen in die Bestimmung des „aufgeklärten“ Anteils neben den Mittelwertsdifferenzen auch Regressionskoeffizienten ein. In unserem Fall beschreiben diese die Beiträge der einzelnen Merkmale zur Vorhersage der Lesekompetenz in PISA 2009. Wie immer bei Regressionsanalysen ist dabei zu berücksichtigen, dass die Höhe der Regressionskoeffizienten der einzelnen Prädiktoren unter anderem durch die Auswahl der übrigen Prädiktoren beeinflusst wird. Auch vor diesem Hintergrund sollten die Ergebnisse nicht als kausale Nachweise interpretiert werden, sondern lediglich als einfache Modellrechnungen, die erste Hinweise auf erklärende Merkmale liefern können.

Die eingeschränkte Interpretierbarkeit der Analysen verweist noch einmal darauf, wie schwierig es ist, anhand von Querschnittsdaten die Effekte von Reformbemühungen in einem Bildungssystem von strukturellen Veränderungen über die Zeit zu separieren. Um solche Analysen zu erleichtern, sollten in zukünftigen Erhebungen potenziell wichtige Prozessmerkmale (z.B. Angebote der Sprach- und Leseförderung) über die Zeit in vergleichbarer Weise erfasst werden. In weiteren Untersuchungen wäre ferner zu prüfen, inwieweit sich die Befunde auch für andere Domänen wie Mathematik und Naturwissenschaften nachweisen lassen. In PISA können allerdings die Veränderungen über die Zeit in Mathematik erst ab PISA 2012 und in den Naturwissenschaften erst ab PISA 2015 zuverlässig festgestellt werden. Entsprechende Analysen für diese beiden Kompetenzbereiche werden mit den Daten aus diesen nachfolgenden PISA-Zyklen möglich werden.

Literatur

- Ammermueller, A. (2007). PISA: What makes the difference? Explaining the gap in test scores between Finland and Germany. *Empirical Economics*, 33, 263-287.
- Autorengruppe Bildungsberichterstattung (2012). *Bildung in Deutschland 2012. Ein indikatoren-gestützter Bericht mit einer Analyse zur kulturellen Bildung im Lebenslauf*. Bielefeld: Bertelsmann.
- BaMF = Bundesamt für Migration und Flüchtlinge (2007). *Bundesweites Integrationsprogramm (§ 45 Aufenthaltsgesetz): Feststellung der Sprachförderangebote des Bundes und der Länder. Dokumentation*. Nürnberg: BaMF.

- Baumert, J., Stanat, P., & Demmrich, A. (2001). PISA 2000: Untersuchungsgegenstand, theoretische Grundlagen und Durchführung der Studie. In J. Baumert, E. Klieme, M. Neubrand, M. Prenzel, U. Schiefele, W. Schneider, P. Stanat, K.-J. Tillmann & M. Weiss (Hrsg.), *PISA 2000: Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich* (S. 15-68). Opladen: Leske + Budrich.
- Baumert, J., Stanat, P., & Watermann, R. (2006). Schulstruktur und die Entstehung differenzieller Lern- und Entwicklungsmilieus. In Dies. (Hrsg.), *Herkunftsbedingte Disparitäten im Bildungswesen: Differenzielle Bildungsprozesse und Probleme der Verteilungsgerechtigkeit. Vertiefende Analysen im Rahmen von PISA 2000* (S. 95-188). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Becker, M., Lüdtke, O., Trautwein, U., & Baumert, J. (2006). Leistungszuwachs in Mathematik: Evidenz für einen Schereneffekt im mehrgliedrigen Schulsystem. *Zeitschrift für Pädagogische Psychologie*, 20, 233-242.
- Blinder, A. S. (1973). Wage Discrimination: Reduced form and structural estimates. *The Journal of Human Resources*, 8, 436-455.
- Cattaneo, M. A., & Wolter, S. (2012). *Migration Policy Can Boost PISA Results – Findings from a Natural Experiment* (SKBF Staff Paper 7). Bern: Schweizerische Koordinationsstelle für Bildungsforschung.
- Ehmke, T., & Jude, N. (2010). Soziale Herkunft und Kompetenzerwerb. In E. Klieme, C. Artelt, J. Hartig, N. Jude, O. Köller, M. Prenzel, W. Schneider & P. Stanat (Hrsg.), *PISA 2009. Bilanz nach einem Jahrzehnt* (S. 231-254). Münster: Waxmann.
- Hanushek, E. A., & Wössmann, L. (2010). *The economics of international differences in educational achievement* (NBER Working Papers 15949, National Bureau of Economic Research, Cambridge). <http://www.nber.org/papers/w15949.pdf> [03.11.2011].
- Holland, P. W., & Dorans, N. J. (2006). Linking and Equating. In R. L. Brennan (Hrsg.), *Educational Measurement* (S. 187-220). Westport: ACE/Praeger.
- Jude, N., & Klieme, E. (2010). Das Programme für International Student Assessment (PISA). In E. Klieme, C. Artelt, J. Hartig, N. Jude, O. Köller, M. Prenzel, W. Schneider & P. Stanat (Hrsg.), *PISA 2009. Bilanz nach einem Jahrzehnt* (S. 11-22). Münster: Waxmann.
- Klieme, E., Artelt, C., Hartig, J., Jude, N., Köller, O., Prenzel, M., Schneider, W., & Stanat, P. (2010). *PISA 2009. Bilanz nach einem Jahrzehnt*. Münster: Waxmann.
- Klieme, E., Jude, N., Baumert, J., & Prenzel, M. (2010). PISA 2000-2009: Bilanz der Veränderungen im Schulsystem. In E. Klieme, C. Artelt, J. Hartig, N. Jude, O. Köller, M. Prenzel, W. Schneider & P. Stanat (Hrsg.), *PISA 2009. Bilanz nach einem Jahrzehnt* (S. 277-296). Münster: Waxmann.
- Konsortium Bildungsberichterstattung (2006). *Bildung in Deutschland. Ein indikatorengestützter Bericht mit einer Analyse zu Bildung und Migration*. Bielefeld: Bertelsmann.
- Naumann, J., Artelt, C., Schneider, W., & Stanat, P. (2010). Lesekompetenz von PISA 2000 bis PISA 2009. In E. Klieme, C. Artelt, J. Hartig, N. Jude, O. Köller, M. Prenzel, W. Schneider & P. Stanat (Hrsg.), *PISA 2009. Bilanz nach einem Jahrzehnt* (S. 23-72). Münster: Waxmann.
- Oaxaca, R. L. (1973). Male-Female Wage Differentials in Urban Labor Markets. *International Economic Review*, 14(3), 693-709.
- OECD (2004). *Lernen für die Welt von morgen. Erste Ergebnisse von PISA 2003*. Paris: OECD.
- OECD (2009). *PISA 2009 assessment framework. Key competencies in reading, mathematics and science*. Paris: OECD.
- OECD (2010). *PISA 2009 Results: Learning Trends: Changes in Student Performance since 2000* (Volume V). Paris: OECD.
- Paetsch, J., Wolf, K., & Stanat, P. (2010). *Förderung von Kindern und Jugendlichen aus Zuwandererfamilien. Expertise für den Expertenrat „Herkunft und Bildungserfolg“ im Auftrag des Ministeriums für Kultur, Jugend und Sport des Landes Baden-Württemberg* (unveröffentlichtes Manuskript).

- Redder, A., Schwippert, K., Hasselhorn, M., Forscher, S., Fickermann, D., & Ehlich, K. (2010). *Grundzüge eines nationalen Forschungsprogramms zu Sprachdiagnostik und Sprachförderung* (ZUSE-Diskussionspapier Nr. 1). Hamburg: Hamburger Zentrum zur Unterstützung der wissenschaftlichen Begleitung und Erforschung schulischer Entwicklungsprozesse (ZUSE).
- Stanat, P., Rauch, D., & Segeritz, M. (2010). Schülerinnen und Schüler mit Migrationshintergrund. In E. Klieme, C. Artelt, J. Hartig, N. Jude, O. Köller, M. Prenzel, W. Schneider & P. Stanat (Hrsg.), *PISA 2009. Bilanz nach einem Jahrzehnt* (S. 200-230). Münster: Waxmann.
- Zhang, L., & Lee, K. A. (2011). Decomposing achievement gaps among OECD countries. *Asia Pacific Education Review*, 12, 463-474.

Anschrift der Autoren/der Autorin

Prof. Dr. Timo Ehmke, Leuphana Universität Lüneburg, Fakultät I Institut für Bildungswissenschaft, Scharnhorststraße 1, 21335 Lüneburg, Deutschland
E-Mail: tehmke@leuphana.de

Prof. Dr. Eckhard Klieme, Deutsches Institut für Internationale Pädagogische Forschung (DIPF), Schloßstraße 29, 60486 Frankfurt am Main, Deutschland
E-Mail: klieme@dipf.de

Prof. Dr. Petra Stanat, Humboldt-Universität zu Berlin, Institut zur Qualitätsentwicklung im Bildungswesen (IQB), Unter den Linden 6, 10099 Berlin, Deutschland
E-Mail: iqboffice@iqb.hu-berlin.de

Dorothea Mildner/Jan Hochweber/Andreas Frey

Vergleichende Analysen der Kompetenzen von Fünfzehnjährigen und Neuntklässlern in den deutschen PISA-Erhebungen 2003 bis 2009

1. Einleitung

Das *Programme for International Student Assessment* (PISA) wurde im Jahr 2009 zum vierten Mal durchgeführt. Die bei den ersten PISA-Erhebungen zentralen querschnittlichen Betrachtungen können somit inzwischen durch differenzierte Trendanalysen in ihrer Aussagekraft bedeutend erweitert werden. Bei PISA 2009 bildeten Trendanalysen bereits einen zentralen Bestandteil der regulären PISA-Berichterstattung (vgl. Klieme, Jude, Baumert & Prenzel, 2010; OECD, 2010).

Trendanalysen verfolgen zwei grundlegende Ziele. Das erste Ziel besteht im sogenannten *Benchmarking*. Beim Benchmarking werden Veränderungen der Kompetenzstände von Schülerinnen und Schülern zwischen den an PISA teilnehmenden Staaten (beziehungsweise Bildungssystemen) verglichen. Hierbei kann beispielsweise analysiert werden, inwieweit Veränderungen der mittleren Kompetenzen und gegebenenfalls deren Streuungen in verschiedenen Staaten gleich oder unterschiedlich ausfallen. Allerdings sind Trendanalysen für einzelne Teilnehmerstaaten nicht erst durch den Vergleich mit anderen Staaten interessant. Das zweite generelle Ziel, das *Monitoring*, betrachtet deshalb Veränderungen von Kompetenzverteilungen innerhalb eines Staates über die Zeit. Hierbei werden die Veränderungen häufig mit als relevant erachteten Bedingungsfaktoren für den Kompetenzerwerb in Beziehung gesetzt. So können Trendanalysen zum Beispiel Hinweise auf den Erfolg von Reformmaßnahmen im Bildungsbereich geben. Zudem gestatten die bei PISA verwendeten Kompetenzskalen, dass Messwerte kriteriumsorientiert und nicht nur normorientiert interpretiert werden können. Während bei normorientierten Testwertinterpretationen nur relative Vergleiche mit Referenzgruppen wie anderen Staaten möglich sind, erlauben kriteriumsorientierte Testwertinterpretationen Aussagen darüber, was Schülerinnen und Schüler mit einer gegebenen Wahrscheinlichkeit inhaltlich wissen und können (vgl. Herzberg & Frey, 2011). Damit gewinnen Trendaussagen auch jenseits des Staatenvergleichs an Aussagekraft, etwa dahingehend, inwieweit es einem Teilnehmerstaat gelungen ist, den Anteil sehr schwacher Leserinnen und Leser über einen gewissen Zeitraum hinweg zu reduzieren (z.B. Naumann, Artelt, Schneider & Stanat, 2010).

Eine nationale Perspektive auf Trends ist für Deutschland aus einem weiteren Grund von besonderem Interesse. In Deutschland wird bereits seit der ersten PISA-Erhebung im Jahr 2000 neben der altersbasierten Stichprobe von fünfzehnjährigen Schülerinnen und Schülern eine jahrgangsstufenbasierte Stichprobe von Neuntklässlerinnen und Neuntklässlern erhoben (vgl. Abschnitt 2). Da jahrgangsstufenbasierte Stichproben je-

weils nur in einigen Staaten untersucht wurden, bezog sich die internationale Trendberichterstattung bislang ausschließlich auf Trends für die fünfzehnjährigen Schülerinnen und Schüler. Eine Betrachtung von Trends für die Schülerinnen und Schüler der neunten Jahrgangsstufe ermöglicht darüber hinaus jedoch sehr interessante und neue Erkenntnisse. Es sind keineswegs stets parallele Kompetenzverläufe über die Zeit für Fünfzehnjährige und Neuntklässlerinnen und Neuntklässler zu erwarten. So könnten beispielsweise Veränderungen der Bildungsbeteiligung, des mittleren Alters von Neuntklässlerinnen und Neuntklässlern, der Anteile von Jugendlichen, die eine Klasse wiederholt haben, oder spezifische Veränderungen der Kompetenzen von Subpopulationen zu Differenzen zwischen den Trendverläufen von Fünfzehnjährigen und Neuntklässlerinnen und Neuntklässlern führen. Derartige Differenzen sind für ein umfassendes Verständnis der Funktionsweise von Bildungssystemen hoch relevant.

Im vorliegenden Beitrag geht es entsprechend darum, die Trendberichterstattung in Deutschland, die sich bisher auf die Betrachtung der Trends bei den fünfzehnjährigen Schülerinnen und Schülern beschränkt, um die Betrachtung von Trends für die neunte Jahrgangsstufe zu erweitern. Dabei werden die Trends der Kompetenzen im Lesen, in der Mathematik und in den Naturwissenschaften zwischen PISA 2003 und PISA 2009 in den beiden Stichproben verglichen. Darüber hinaus soll exemplarisch versucht werden, mögliche Ursachen für unterschiedliche Trends in den Stichproben zu identifizieren. Dabei beschränken wir uns auf die Betrachtung der Veränderung in der Jahrgangsstufenverteilung fünfzehnjähriger Schülerinnen und Schüler.

Im Folgenden wird zunächst dargestellt, worin sich eine lebensalterbasierte Populationsdefinition – wie sie bei der internationalen PISA-Erhebung vorgenommen wird – und eine jahrgangsstufenbasierte Populationsdefinition – wie sie der nationalen Stichprobe der neunten Klassen in Deutschland zugrunde liegt – unterscheiden. Um zu prüfen, inwieweit die Jahrgangsstufenverteilung fünfzehnjähriger Schülerinnen und Schüler unterschiedliche Trends in den beiden PISA-Stichproben erklären könnte, wird die Veränderung dieser Verteilung zwischen PISA 2003 und PISA 2009 betrachtet. Auf die daran anschließende genauere Erläuterung der Zielsetzung dieses Beitrags folgt im Methodenteil eine Darstellung der analysierten PISA-Stichproben zwischen PISA 2003 und PISA 2009, der Datengrundlage sowie der statistischen Verfahren, die für die Vergleiche zwischen den Erhebungszeitpunkten eingesetzt wurden. Im Anschluss werden die Ergebnisse der Trendanalysen für die Fünfzehnjährigen und die Neuntklässlerinnen und Neuntklässler berichtet und mögliche Beziehungen zu Veränderungen in der Verteilung der Fünfzehnjährigen auf die Jahrgangsstufen untersucht. Abschließend erfolgen ein Fazit der Ergebnisse sowie ein kurzer Ausblick.

2. Lebensalterbasierte und jahrgangsstufenbasierte Stichproben

Eine zentrale Aufgabe für eine internationale Schulleistungsstudie wie PISA ist es, verlässliche internationale Vergleiche von Kompetenzständen zu gewährleisten (Baumert & Stanat, 2006). Ein grundlegender Aspekt ist dabei die Definition der interessierenden

Zielpopulation, auf die die Ergebnisse generalisiert werden sollen. Angesichts der Organisation der meisten Bildungssysteme liegt es intuitiv nahe, eine jahrgangsstufenbasierte Definition der Zielpopulation vorzunehmen. Dabei werden in jedem Staat Schülerinnen und Schüler untersucht, die eine bestimmte Jahrgangsstufe besuchen. Allerdings führt gerade diese Vorgehensweise oftmals zu erheblichen Schwierigkeiten. International vergleichbare Jahrgangsstufen sind häufig schwierig zu identifizieren, da sich Staaten in der Art und dem Ausmaß an vorschulischer Bildung und Erziehung, im Schuleintrittsalter sowie in der institutionellen Struktur ihrer nationalen Bildungssysteme teilweise erheblich unterscheiden.

Folglich definieren internationale Schulleistungsvergleiche typischerweise ihre Zielpopulation über ein bestimmtes Lebensalter. Einige internationale Studien (z.B. Progress in International Reading Literacy Study – PIRLS; Third International Mathematics and Science Study – TIMSS) haben dabei ihre Zielpopulation auf Basis von Jahrgängen definiert, die eine maximale Abdeckung einer bestimmten Alterskohorte bieten. Ein Nachteil dieses Ansatzes liegt darin, dass geringe Variationen in der Altersverteilung zwischen den Jahrgangsstufen zur Auswahl unterschiedlicher Zieljahrgangsstufen in den Teilnehmerstaaten oder gar zwischen Bildungssystemen innerhalb von Staaten führen können. Kompetenzvergleiche können auch dadurch erschwert werden, wenn nicht alle Schülerinnen und Schüler des interessierenden Alters in den jahrgangsbasierten Stichproben repräsentiert sind und sich die nicht repräsentierten Schülerinnen und Schüler typischerweise in der nächsthöheren Jahrgangsstufe in einigen Staaten und in der nächstniedrigeren Jahrgangsstufe in anderen Staaten befinden. Dieser Umstand würde im ersten Fall Schülerinnen und Schüler auf potenziell höheren Kompetenzstufen, im zweiten Fall solche auf potenziell niedrigeren Kompetenzstufen ausschließen.

Um derartige Probleme zu vermeiden, wird bei PISA in Übereinstimmung mit dem zugrunde liegenden *Literacy*-Konzept¹ (z.B. Baumert, Stanat & Demmrich, 2001) eine lebensalterbasierte Definition der Zielpopulation genutzt. Es wird damit eine Definition verwendet, die in geringerem Ausmaß an die institutionellen Strukturen nationaler Bildungssysteme gebunden ist: Unabhängig von der besuchten Jahrgangsstufe oder der Art der Bildungseinrichtung werden Schülerinnen und Schüler untersucht, die zu Beginn des Testzeitraumes fünfzehn Jahre alt sind. Mit dieser Populationsdefinition macht PISA Aussagen über das Wissen und die Fähigkeiten einer Gruppe von Individuen, die innerhalb einer vergleichbaren Referenzperiode geboren wurden, die aber durchaus unterschiedliche Bildungserfahrungen innerhalb und außerhalb der Schule gemacht haben können. Bei PISA werden dieses Wissen und diese Fähigkeiten als Bildungserträge im Alter von fünfzehn Jahren interpretiert. In Abhängigkeit von den nationalen Regelungen der Staaten zum Schuleintritt, zur Schulwahl und zum Schulsystem können diese Schü-

1 Ausgangspunkt der Rahmenkonzeption von PISA ist eine Vorstellung von Grundbildung (engl. literacy), bei der es um die Fähigkeit geht, Kenntnisse und Fertigkeiten in einer Vielzahl von Situationen anzuwenden und zu analysieren, logisch zu denken und in effektiver Weise zu kommunizieren. Innerhalb der bei PISA untersuchten Domänen (Lesen, Mathematik, Naturwissenschaften) beruht Literacy auf bereichsspezifischem Wissen.

lerinnen und Schüler je nach Bildungssystem und Bildungszweig auf unterschiedliche Jahrgangsstufen verteilt sein. Solche Verschiedenheiten müssen berücksichtigt werden, wenn PISA-Ergebnisse zwischen Teilnehmerstaaten verglichen werden. Sind die Kompetenzwerte in den bei PISA untersuchten Bereichen Lesen, Mathematik und Naturwissenschaften in einem Staat signifikant höher als in einem anderen Staat, so kann daraus nicht automatisch geschlossen werden, dass die Schulen oder bestimmte Teile des Bildungssystems im ersten Staat effektiver sind als im zweiten. Es kann jedoch geschlossen werden, dass der kumulative Bildungsertrag bis zum Alter von fünfzehn Jahren im ersten Staat höher ist als im zweiten.

Bei einer altersbasierten Definition der Zielpopulation muss allerdings in Kauf genommen werden, dass die Möglichkeiten eingeschränkt sind, gefundene Unterschiede auf institutionelle Bedingungen zurückzuführen. So spielt die Zuordnung der Schülerinnen und Schüler zu Schulklassen oder Kursen – welche für die Institution Schule ein zentrales Charakteristikum und insbesondere für die Organisation von Unterricht von überragender Bedeutung ist – für die Stichprobengewinnung und letztlich für die Analysen im Rahmen der internationalen PISA-Erhebung keine Rolle. Um Staaten entgegenzukommen, die jahrgangsbasierte Ergebnisse im Rahmen nationaler Analysen anstreben, bietet das Design von PISA eine Option zur nationalen Stichprobenerweiterung, welche die altersbasierte Stichprobe um eine jahrgangsbasierte Stichprobe ergänzt. Von dieser Möglichkeit wurde in Deutschland bereits bei PISA 2000 Gebrauch gemacht. Allerdings wurden dort keine vollständigen Klassen als Stichprobeneinheiten gezogen, sondern zusätzlich zu den bereits in der Stichprobe der Fünfzehnjährigen enthaltenen Schülerinnen und Schülern aus neunten Klassen zehn weitere nicht fünfzehnjährige Neuntklässlerinnen und Neuntklässler zufällig ausgewählt, um so für diese Jahrgangsstufe auch die übrigen in ihr enthaltenen Altersgruppen zu repräsentieren (vgl. Sibberns & Baumert, 2001, S. 511). Seit PISA 2003 wird an den für den internationalen Vergleich ausgewählten Schulen die international verbindliche Stichprobe der Fünfzehnjährigen so ergänzt, dass dort zusätzlich jeweils zwei komplette Klassen der neunten Jahrgangsstufe erhoben werden. Die Ziehung der nationalen Stichproben der neunten Jahrgangsstufe wird somit erst ab PISA 2003 in identischer Weise vorgenommen.

3. Verteilung der Fünfzehnjährigen auf die Jahrgangsstufen

Die wiederholte Erhebung von Kompetenzen bei Fünfzehnjährigen und Neuntklässlerinnen und Neuntklässlern ermöglicht es, Unterschiede in den Trendverläufen zwischen beiden Stichproben zu beschreiben. Darüber hinaus können diese Unterschiede mit Merkmalen der Stichproben in Beziehung gesetzt werden, was im vorliegenden Beitrag exemplarisch für die Verteilung der Fünfzehnjährigen auf die Jahrgangsstufen geschehen soll. Voraussetzungen für einen Erklärungsbeitrag der Jahrgangsstufenverteilung wären, dass (a) sich diese Verteilung im betrachteten Zeitraum in der Zielpopulation der Fünfzehnjährigen wesentlich verändert hat und (b) für den Fall einer solchen

Veränderung bedeutsame Unterschiede in den Kompetenzen der Fünfzehnjährigen zwischen den PISA-Erhebungen zu erwarten sind.

Zumindest zwei Entwicklungen im deutschen Bildungssystem sprechen dafür, dass der Besuch höherer Jahrgangsstufen durch die fünfzehnjährigen Schülerinnen und Schüler seit dem Beginn der PISA-Erhebungen zugenommen hat (Klieme et al., 2010). So hat in den Alterskohorten, aus denen sich die Teilnehmerinnen und Teilnehmer der PISA-Erhebungen rekrutierten, der Anteil vorzeitiger Einschulungen erheblich zugenommen, während der Anteil von Zurückstellungen bei der Einschulung kontinuierlich abgenommen hat. Auch die Anzahl der Klassenwiederholungen ging gemäß der amtlichen Schulstatistik, wenn auch nur im Primarbereich, in dem für die PISA-Teilnehmerinnen und -Teilnehmer relevanten Zeitraum von 1995 bis 2004 (vgl. Klieme et al., 2010) leicht zurück. Die prozentualen Anteile der fünfzehnjährigen Schülerinnen und Schüler in den verschiedenen Jahrgangsstufen in den PISA-Erhebungen 2003 bis 2009 werden in Tabelle 1 berichtet.² Ebenfalls dargestellt sind die Differenzen der Prozentwerte zwischen den einzelnen PISA-Erhebungen.

Jahgangs- stufe	PISA 2003	PISA 2006	PISA 2009	Diff 06–03	Diff 09–06	Diff 09–03
7	1.7	1.6	1.2	–0.1	–0.3	–0.5
8	15.0	12.3	11.0	–2.6	–1.4	–4.0
9	59.9	56.5	54.8	–3.4	–1.8	–5.2
10	23.2	29.3	32.5	6.0	3.3	9.3
11	0.1	0.3	0.4	0.2	0.2	0.3
12			< 0.1		< 0.1	< 0.1

Tab. 1: *Prozentuale Verteilung der Fünfzehnjährigen auf die Jahrgangsstufen in den PISA-Erhebungen 2003 bis 2009 sowie die Differenzen (Diff) der Prozentwerte zwischen den PISA-Erhebungen (gewichtet), Werte gerundet*

Die prozentualen Veränderungen stützen die erwartete deutliche Verschiebung hin zum Besuch höherer Jahrgangsstufen; die Anteile in den Jahrgangsstufen 8 und 9 nehmen zusammen um gut neun Prozent ab, der Anteil in Jahrgangsstufe 10 um nahezu den glei-

2 Es ließe sich einwenden, dass die Ermittlung der Verteilung basierend auf den PISA-Stichproben weniger aussagekräftig sei als der Rückgriff auf die Daten der amtlichen Schulstatistik. Allerdings beruhen Stichprobenziehung und Berechnung der Gewichte der Schülerinnen und Schüler bei PISA auf den jeweils aktuellen zum Zeitpunkt der Stichprobenziehung verfügbaren Angaben der Statistischen Landesämter sowie des Statistischen Bundesamts und der Landesschulämter (Carstensen, Frey, Walter & Knoll, 2007). Entsprechend sind basierend auf den PISA-Stichproben zuverlässige Aussagen über die Gesamtheit der Zielpopulation der fünfzehnjährigen Schülerinnen und Schüler in Deutschland möglich.

chen Wert zu. Die Tendenz ist dabei sowohl zwischen PISA 2003 und PISA 2006 als auch zwischen PISA 2006 und PISA 2009 ersichtlich, wobei die Differenzen im erstgenannten Zeitraum jeweils im Betrag höher ausfallen.

Sind nun angesichts der beschriebenen Veränderungen in der Jahrgangsstufenverteilung bedeutsame positive Veränderungen in den Kompetenzen der Fünfzehnjährigen zu erwarten? Die jeweils besuchte Jahrgangsstufe ist von entscheidender Bedeutung für die Lernzeit, welche Schülerinnen und Schüler in ihrer Schullaufbahn kumulieren. Die Lernzeit, die ihnen zur Verfügung steht, gehört wiederum zu den zentralen Faktoren des individuellen Lernerfolgs (Helmke, 2010; Seidel & Shavelson, 2007). Entsprechend sollten sich Veränderungen in der Verteilung der Fünfzehnjährigen auf die Jahrgangsstufen auch in den mittleren Kompetenzen widerspiegeln. Weiteren Aufschluss geben hier Studien, die den Lernzuwachs innerhalb einzelner Schuljahre am Ende der Sekundarstufe I quantifizieren. In der Studie PISA-I-Plus wurde die Stichprobe der Fünfzehnjährigen aus PISA 2003 ein Jahr später (in der zehnten Jahrgangsstufe) noch einmal mit dem Schwerpunkt Mathematik getestet, um Aufschlüsse über die Kompetenzentwicklung im Verlauf eines Schuljahres zu ermöglichen (Prenzel et al., 2006). Für den grundbildungsorientierten Mathematiktest wurde zwischen der neunten und zehnten Jahrgangsstufe ein Kompetenzzuwachs von Cohens $d = 0.33$ ermittelt, was als deutliche Verbesserung eingestuft werden kann (Ehmke, Blum, Neubrand, Jordan & Ulfig, 2006). Anders stellen sich die Befunde im Bereich des Lesens dar: Während sich in der Hamburger Längsschnittstudie Aspekte der Lernausgangslage und der Lernentwicklung (LAU) zwischen dem Beginn der siebten und dem Ende der achten Jahrgangsstufe jährliche Kompetenzzuwächse von etwa einer Viertel Standardabweichung ergaben (Lehmann, Peek, Gänsfuß & Husfeldt, 2002), resultierte im Lesekompetenztest in der Studie Deutsch Englisch Schülerleistungen International (DESI) kein signifikanter Zuwachs im Laufe der neunten Jahrgangsstufe (Gailberger & Willenberg, 2008).

4. Zielsetzung

Im vorliegenden Beitrag wird untersucht, wie sich die mittleren Kompetenzen der fünfzehnjährigen Schülerinnen und Schüler sowie der Schülerinnen und Schüler in neunten Jahrgangsstufen in den Domänen Lesen, Mathematik und Naturwissenschaften zwischen PISA 2003 und PISA 2009 in Deutschland verändert haben. Bei allen Analysen beschränken wir uns auf die Betrachtung von Trends in den Kompetenzmittelwerten der Schülerinnen und Schüler. Weiterhin beschränkt sich unsere Darstellung auf die PISA-Erhebungen der Jahre 2003 bis 2009 im Lesen und in der Mathematik sowie der Jahre 2006 und 2009 in den Naturwissenschaften. Dies geschieht aus zwei Gründen: Erstens wird erst seit PISA 2003 bei der nationalen Stichprobenerweiterung um die Schülerinnen und Schüler aus neunten Klassen ein vergleichbares Vorgehen angewandt (vgl. Abschnitt 2). Zweitens erfolgt für den Bereich Naturwissenschaften auch international erst beginnend mit PISA 2006 eine Trendberichterstattung, da erst zu diesem Zeitpunkt, als die Naturwissenschaften den Schwerpunktbereich von PISA bildeten, die ausdifferen-

zierte Konzeption für die Erfassung naturwissenschaftlicher Kompetenz vorlag (OECD, 2006).

Über die reine Deskription von Trends in den Stichproben der Fünfzehnjährigen und der Neuntklässlerinnen und Neuntklässler hinaus stellt sich die Frage, welche Ursachen möglichen Unterschieden zwischen den Trends in den Stichproben zugrunde liegen. Als Erklärungsfaktoren bieten sich hierfür insbesondere Veränderungen in strukturellen Merkmalen des Bildungssystems an, welches sich in den zurückliegenden Jahren seit PISA 2000 in verschiedener Hinsicht bedeutend gewandelt hat (siehe Klieme et al., 2010). Wir werden uns hier auf einen potenziell zentralen Erklärungsfaktor, die Veränderung in der Verteilung fünfzehnjähriger Schülerinnen und Schüler auf die Jahrgangsstufen, beschränken und gehen dabei von folgenden Annahmen aus: Die mit dem zunehmenden Besuch höherer Jahrgangsstufen (vgl. Abschnitt 3) einhergehende längere Beschulung sollte sich insgesamt positiv auf die Kompetenzmittelwerte der Fünfzehnjährigen auswirken. Die von den Neuntklässlerinnen und Neuntklässlern besuchte Jahrgangsstufe ist dagegen definitionsgemäß konstant, sodass – bei Berücksichtigung dieses einen Erklärungsfaktors – von einer günstigeren Veränderung der Kompetenzmittelwerte bei den Fünfzehnjährigen verglichen mit den Neuntklässlerinnen und Neuntklässlern auszugehen ist. Aus den in Abschnitt 3 angeführten Studien zum Lernzuwachs lässt sich ableiten, dass positive Veränderungen in den Kompetenzmittelwerten der Fünfzehnjährigen im Lesen vor allem dann zu erwarten sind, wenn ein höherer Anteil von Schülerinnen und Schülern die Jahrgangsstufe 7 beziehungsweise 8 bereits absolviert hat. Abnehmende Anteile in Jahrgangsstufe 9 zugunsten zunehmender Anteile in Jahrgangsstufe 10 sollten weniger bedeutsam sein, da sich im Lesen für Jahrgangsstufe 9 keine Lernzuwächse nachweisen ließen. In der Mathematik hingegen resultierten deutliche Lernzuwächse auch für die neunte Jahrgangsstufe; entsprechend sollten eine Abnahme des Anteils der Fünfzehnjährigen in der neunten Jahrgangsstufe und eine Zunahme des Anteils in der zehnten Jahrgangsstufe mit positiven Veränderungen in den Kompetenzmittelwerten einhergehen. Betrachtet man die Veränderungen in der Jahrgangsstufenverteilung zwischen PISA 2003 und PISA 2009 (vgl. Abschnitt 3), so ergibt sich angesichts der Verringerung der Anteile in den Jahrgangsstufen 8 und 9 und der Zunahme des Anteils in Jahrgangsstufe 10 die Erwartung eines vergleichsweise deutlichen Kompetenzzuwachses im Bereich Mathematik und eines geringer ausgeprägten Kompetenzzuwachses im Bereich Lesen. Für die Naturwissenschaften lassen sich aus bisherigen Studien keine eindeutigen Vorhersagen ableiten. Allerdings zeigen sich zwischen den relevanten PISA-Erhebungen 2006 und 2009 nur vergleichsweise geringe Veränderungen hin zum Besuch höherer Jahrgangsstufen. Entsprechend wäre im Bereich Naturwissenschaften eine allenfalls schwache Verbesserung des Kompetenzmittelwerts zu erwarten.

5. Methode

Im Folgenden werden zunächst die analysierten Stichproben aus den PISA-Erhebungen 2003 bis 2009 dargestellt, bevor die statistischen Verfahren kurz erläutert werden, die für die Vergleiche zwischen den Erhebungszeitpunkten eingesetzt wurden.

5.1 Stichprobenbeschreibung

Insgesamt wurden sechs Stichproben analysiert. Für die PISA-Erhebungen 2003, 2006 und 2009 wurden jeweils die internationale Stichprobe der Fünfzehnjährigen sowie die nationale Stichprobenerweiterung der neunten Jahrgangsstufe in Deutschland einbezogen.³ Nachfolgend werden die analysierten Stichproben einzeln beschrieben.

PISA 2003

Internationale Stichprobe der Fünfzehnjährigen

Zwischen März und Mai 2003 nahmen am internationalen Vergleich 216 Schulen in Deutschland teil. Innerhalb dieser Schulen wurden Zufallsstichproben von fünfzehnjährigen Schülerinnen und Schülern gezogen. An den Sonder- und Förderschulen sowie an den Beruflichen Schulen umfasste die Stichprobe alle Fünfzehnjährigen, in den anderen Schulen wurden jeweils 25 Fünfzehnjährige in die Stichprobe einbezogen. Insgesamt umfasst die für den internationalen Vergleich im Rahmen von PISA 2003 gezogene Stichprobe 4660 fünfzehnjährige Schülerinnen und Schüler.

Nationale Stichprobe der neunten Jahrgangsstufe

Bei PISA 2003 wurden in den am internationalen Vergleich teilnehmenden allgemeinbildenden Schulen zusätzlich zwei vollständige neunte Klassen pro Schule zufällig gezogen. Die Schülerinnen und Schüler der Stichprobe bearbeiteten am ersten Testtag – wie die Fünfzehnjährigen – den internationalen PISA-Test und an einem zweiten Testtag zusätzliche nationale (stärker curricular orientierte) Tests für Mathematik und Naturwissenschaften sowie einen ergänzenden nationalen Fragebogen. Die nationale PISA-2003-Stichprobe umfasst insgesamt 8559 Neuntklässlerinnen und Neuntklässler an 198 Schulen. Die Differenz in der Anzahl der Schulen zur internationalen Stichprobe ergibt sich durch die teilnehmenden neun Sonder- und Förderschulen sowie neun Beruflichen Schulen.

PISA 2006

Internationale Stichprobe der Fünfzehnjährigen

In Deutschland wurden zwischen April und Mai 2006 für den internationalen Vergleich Schülerinnen und Schüler aus 226 Schulen getestet. Den internationalen Vorgaben folgend wurden pro Schule 25 fünfzehnjährige Schülerinnen und Schüler aller Jahrgangs-

3 Anders als die PISA-Stichproben der Fünfzehnjährigen enthalten die PISA-Stichproben der Neuntklässlerinnen und Neuntklässler keine Schülerinnen und Schüler aus Sonder- und Förderschulen sowie Beruflichen Schulen. Da bei der Trendberichterstattung im Rahmen von PISA jeweils die gesamten Stichproben der Fünfzehnjährigen zugrunde gelegt werden, wird auch im Folgenden die gesamte Stichprobe der Fünfzehnjährigen inklusive dieser Schülerinnen und Schüler herangezogen (Anteile der Schülerinnen und Schüler in Sonder- und Förderschulen sowie Beruflichen Schulen: PISA 2003: 6.3%, PISA 2006: 7.1%, PISA 2009: 7.7%).

stufen zufällig gezogen. Die Stichprobe umfasst 4891 fünfzehnjährige Schülerinnen und Schüler.

Nationale Stichprobe der neunten Jahrgangsstufe

Die internationale Stichprobe wurde bei PISA 2006 in Deutschland durch die zufällige Ziehung von zwei kompletten Klassen der neunten Jahrgangsstufe je für den internationalen Vergleich gezogener Schule ergänzt. Die Schülerinnen und Schüler bearbeiteten am ersten Testtag den internationalen PISA-Test und an einem zweiten Testtag Mathematikaufgaben zur Überprüfung des Erreichens der Bildungsstandards in Mathematik für den Mittleren Schulabschluss (KMK, 2004a). Die Stichprobe umfasst 9577 Schülerinnen und Schüler an 203 Schulen. Die Differenz in der Anzahl der Schulen zur internationalen Stichprobe ergibt sich durch die teilnehmenden Sonder- und Förderschulen sowie Beruflichen Schulen.

PISA 2009

Internationale Stichprobe der Fünfzehnjährigen

PISA 2009 wurde zwischen April und Mai 2009 an Schulen aller Schularten in Deutschland durchgeführt. Die Stichprobe umfasste gemäß den internationalen Vorgaben pro Schule 25 fünfzehnjährige Schülerinnen und Schüler aller Jahrgangsstufen an insgesamt 226 Schulen in Deutschland. In die Stichprobe gehen Daten von 4979 Schülerinnen und Schülern ein.

Nationale Stichprobe der neunten Jahrgangsstufe

Darüber hinaus wurden an den an PISA 2009 teilnehmenden allgemeinbildenden Schulen pro Schule zusätzlich zwei komplette neunte Klassen getestet. Diese nahmen am ersten Testtag am internationalen PISA-Test teil und an einem zweiten Testtag an den Tests des ersten Ländervergleichs zur Überprüfung der Bildungsstandards für den Mittleren Schulabschluss in den Fächern Deutsch, Englisch und Französisch (KMK, 2004b, c; Köller, Knigge & Tesch, 2010). Die Stichprobe umfasst 9461 Schülerinnen und Schüler an 202 Schulen. Die Differenz in der Anzahl der Schulen zur internationalen Stichprobe ergibt sich durch die teilnehmenden 13 Sonder- und Förderschulen sowie elf Beruflichen Schulen.

5.2 Datengrundlage

Die Datengrundlage der analysierten Stichproben (vgl. Abschnitt 5.1) bilden im Falle der Fünfzehnjährigen Kompetenzwerte für Deutschland aus der internationalen PISA-Skalierung, die auch Grundlage der Berichterlegung sind (z.B. Klieme et al., 2010; OECD, 2010). Bei den Stichproben der neunten Jahrgangsstufe beruhen die Kompetenzwerte auf nationalen Skalierungen, die unter Federführung der jeweils mit der Projektkoordination in Deutschland beauftragten Institute durchgeführt wurden. Für PISA 2003 und

PISA 2006 war dies das Leibniz-Institut für die Pädagogik der Naturwissenschaften und Mathematik (IPN) in Kiel und für PISA 2009 das Deutsche Institut für Internationale Pädagogische Forschung (DIPF) in Frankfurt am Main. Das Vorgehen bei der nationalen Skalierung orientiert sich am internationalen Vorgehen (OECD, 2011) und unterscheidet sich davon im Wesentlichen durch die Berücksichtigung weiterer Variablen aus dem Fragebogen für Schülerinnen und Schüler, die als zusätzliche Hintergrundinformationen in die Skalierung mit einfließen (siehe z.B. Carstensen et al., 2007). Die nationale Skalierung bei PISA 2009 wurde zur Gewährleistung von Vergleichbarkeit mit den früheren PISA-Erhebungen in enger Abstimmung zwischen DIPF und IPN durchgeführt. Eine Vergleichbarkeit der verwendeten Kompetenzdaten zwischen den Stichproben der Fünfzehnjährigen und der neunten Jahrgangsstufe ist ebenfalls gegeben, da bei der nationalen Skalierung (a) das internationale Hintergrundmodell übernommen und im Weiteren ausschließlich ergänzt wurde und (b) die internationalen Itemschwierigkeitsparameter verwendet wurden. Zudem wurde dasselbe mehrdimensionale Skalierungsmodell wie in der internationalen Skalierung herangezogen.

In den im Folgenden durchgeführten Analysen wird mit den *Plausible Values* (PVs; von Davier, Gonzalez & Mislevy, 2009) als Schätzern der Personenfähigkeit gearbeitet, die in den internationalen und nationalen Skalierungen bei PISA 2003, PISA 2006 und PISA 2009 gewonnen wurden. Die PVs erlauben eine unverzerrte Schätzung der Kompetenzmittelwerte und Kompetenzstreuungen sowie der Beziehungen der Kompetenzen zu weiteren Variablen, die – wie die im vorliegenden Beitrag analysierten Variablen – im Hintergrundmodell berücksichtigt wurden. Infolge der Verwendung von PVs lagen in den Stichproben jeweils für alle Schülerinnen und Schüler Kompetenzwerte für alle Domänen (Lesen, Mathematik, Naturwissenschaften) vor, das heißt auch für jene Schülerinnen und Schüler, die aufgrund des Studiendesigns nicht alle Domänen bearbeitet haben.

Gebhardt und Adams (2007) diskutieren drei verschiedene Methoden der Trendskalierung mit PISA-Daten. Als „originale Trends“ werden Veränderungen in den Kompetenzindikatoren bezeichnet, die in den PISA-Erhebungszyklen zur querschnittlichen Berichterstattung konstruiert wurden. Diese originalen Trends basieren auf gemeinsamen Skalierungen der Aufgabenantworten aus den teilnehmenden OECD-Staaten und werden von der OECD im Rahmen ihrer Trendberichterstattung (z.B. OECD, 2010) und auch in der vorliegenden Arbeit verwendet. Unter „marginalen Trends“ werden Veränderungen verstanden, die auf eigenen Skalierungen der Aufgabenantworten für jeden Staat beruhen. Die Aufgabenschwierigkeiten werden für jeden Staat separat geschätzt und es resultiert eine staatenpezifische Kompetenzskala. Sind einzelne Aufgaben oder Aufgabengruppen in einem Staat schwerer oder leichter als in der Gesamtheit der teilnehmenden Staaten, werden diese Unterschiede in den staatenpezifischen Aufgabenschwierigkeitsparametern berücksichtigt. „Bedingte Trends“ sind marginale Trends, bei denen die Veränderungen zwischen den Erhebungszyklen zusätzlich um mögliche Unterschiede in den Stichprobenzusammensetzungen korrigiert werden. In Deutschland wurden im Rahmen von PISA 2006 für längsschnittliche Fragestellungen Ergebnisse einer eigens dafür vorgenommenen marginalen Skalierung berichtet (Carstensen, Prenzel & Baumert, 2008). Es konnte gezeigt werden, dass die originalen und margina-

len Trends für den Bereich Lesen in Deutschland voneinander abweichen und die originalen Trends die Kompetenzen der Schülerinnen und Schüler für PISA 2003 und PISA 2006 überschätzen. Dieser Befund wird in Hinblick auf die Ergebnisse dieses Beitrags in Abschnitt 7 noch einmal aufgegriffen.

Die Zielpopulation der internationalen PISA-Stichprobe in Deutschland bilden alle Schülerinnen und Schüler, die zum Zeitpunkt der Erhebung fünfzehn Jahre alt sind. Die Auswahl der Schulen und der teilnehmenden Schülerinnen und Schüler erfolgt im Rahmen einer zweifach stratifizierten Stichprobenziehung (z.B. Sibberns & Baumert, 2001). Die Wahrscheinlichkeit, in die Stichprobe zu gelangen, ist als Folge dieses Vorgehens nicht für alle Schülerinnen und Schüler gleich. Um repräsentative Aussagen über die Grundgesamtheit der Fünfzehnjährigen zu ermöglichen, werden für alle Schülerinnen und Schüler Gewichte ermittelt (vgl. OECD, 2011). Diese werden zentral durch das internationale PISA-Konsortium berechnet und durch das deutsche Konsortium verifiziert. Nach einem vergleichbaren Vorgehen werden auch für die nationalen Stichproben der neunten Jahrgangsstufe Gewichte entwickelt, die es ermöglichen, auf die Zielpopulation der Neuntklässlerinnen und Neuntklässler in Deutschland zu schließen. Für die im Weiteren dargestellten Analysen werden die finalen Gewichte der Schülerinnen und Schüler (*final student weights*) verwendet (eine Ausnahme wird in Abschnitt 6.2 erläutert).

5.3 Analysen

Bei den nachstehend berichteten Analysen beschränken wir uns auf die Betrachtung von Trends in den Kompetenzmittelwerten der Schülerinnen und Schüler. Mittelwerte und zugehörige Standardfehler werden gemäß dem bei PISA international üblichen Vorgehen errechnet (vgl. OECD, 2011). Unterschiede in den Kompetenzmittelwerten zwischen den verschiedenen Erhebungszeitpunkten werden anhand der *z*-Verteilung auf Signifikanz geprüft. Dabei muss zusätzlich zum Standardfehler der Kompetenzmittelwerte, der sich aus dem Standardmessfehler und dem Sampling-Fehler zusammensetzt, der *Linking-Fehler* berücksichtigt werden. Um die Kompetenzverteilungen verschiedener PISA-Erhebungen aufeinander beziehen zu können, werden die Itemschwierigkeiten für die einzelnen Kompetenzbereiche jeweils an der Berichtsskala jener Erhebung verankert, bei der diese Kompetenz erstmals den inhaltlichen Schwerpunkt darstellte (*Linking*). Das Ausmaß der Variation in den Schwierigkeiten der für das Linking verwendeten Items zwischen den Erhebungen wird durch den Linking-Fehler quantifiziert. Der Linking-Fehler reflektiert somit die statistische Unsicherheit, die auf die Verbindung von Tests verschiedener Erhebungszeitpunkte zurückgeht.

Der Standardfehler der Mittelwertedifferenz zweier Erhebungszeitpunkte berechnet sich nach folgender Formel (exemplarisch für den Vergleich der beiden Erhebungszeitpunkte PISA 2003 und PISA 2009; vgl. OECD, 2011, S. 146):

$$SE = \sqrt{\sigma_{\hat{\mu}_{2003}}^2 + \sigma_{\hat{\mu}_{2009}}^2 + \sigma_{\hat{\mu}_{\text{Linking-Fehler}}}^2}$$

Die Durchführung der berichteten Signifikanztests zur Beurteilung von Mittelwertunterschieden zwischen je zwei Erhebungszeitpunkten erfolgte unter Verwendung der international gültigen Linking-Fehler. Da die bei PISA administrierten Tests für die Stichprobe der Fünfzehnjährigen und die Stichprobe der neunten Jahrgangsstufe jeweils identisch sind und die Skalierung der Stichprobe der neunten Jahrgangsstufe auf Grundlage der international für die Fünfzehnjährigen geschätzten Itemschwierigkeiten (vgl. Abschnitt 5.2) durchgeführt wird, sind für beide Stichproben dieselben Linking-Fehler zu verwenden. Tabelle 2 enthält die Linking-Fehler für die PISA-Erhebungen 2003 bis 2009.

	Linking-Fehler auf der PISA-Skala
Gesamtskala Lesen 2003/2006	5.307
Gesamtskala Lesen 2003/2009	4.088
Gesamtskala Lesen 2006/2009	4.069
Gesamtskala Mathematik 2003/2006	1.382
Gesamtskala Mathematik 2003/2009	1.990
Gesamtskala Mathematik 2006/2009	1.333
Gesamtskala Naturwissenschaften 2006/2009	2.566

Tab. 2: Linking-Fehler für die PISA-Erhebungen 2003 bis 2009

6. Ergebnisse

Zunächst werden die Veränderungen in den mittleren Kompetenzen in den Domänen Lesen und Mathematik zwischen PISA 2003 und PISA 2009 sowie in den Naturwissenschaften zwischen PISA 2006 und PISA 2009 in den Stichproben der Fünfzehnjährigen und der Neuntklässlerinnen und Neuntklässler dargestellt. Anschließend werden Veränderungen in der Verteilung der Fünfzehnjährigen auf die Jahrgangsstufen berichtet und mit den Trends in Beziehung gesetzt.

6.1 Veränderungen der mittleren Kompetenzen in den Stichproben

Lesekompetenz

Die Verläufe der mittleren Lesekompetenz über die Erhebungszeitpunkte (siehe Abbildung 1) unterscheiden sich zwischen den Stichproben der Fünfzehnjährigen und der neunten Jahrgangsstufe nicht auffällig voneinander. Die mittlere Lesekompetenz der Fünfzehnjährigen in Deutschland lag bei PISA 2003 bei 491 Punkten ($SE = 3.4$ Punkte), bei PISA 2006 bei 495 Punkten ($SE = 4.4$ Punkte) und bei PISA 2009 bei 497 Punkten

($SE = 2.7$ Punkte). Die mittlere Lesekompetenz der Fünfzehnjährigen stieg demzufolge über die Erhebungszeitpunkte zumindest deskriptiv kontinuierlich leicht an.

In der Stichprobe der neunten Jahrgangsstufe lag die mittlere Lesekompetenz bei PISA 2003 bei 492 Punkten ($SE = 1.9$ Punkte), bei PISA 2006 wurden 498 Punkte ($SE = 3.5$ Punkte) und bei PISA 2009 494 Punkte ($SE = 2.2$ Punkte) erreicht. Die mittlere Lesekompetenz in der neunten Jahrgangsstufe erreichte also deskriptiv bei PISA 2006 ihre höchste Ausprägung innerhalb des analysierten Zeitraums.

Die größten Differenzen in der Lesekompetenz finden sich mit jeweils sechs Punkten⁴ bei den Fünfzehnjährigen zwischen PISA 2009 und PISA 2003 und für die neunte Jahrgangsstufe zwischen PISA 2006 und PISA 2003. Diese Differenzen sind jedoch – ebenso wie alle anderen Differenzen zwischen den Erhebungszeitpunkten in den beiden Stichproben – statistisch nicht bedeutsam ($p \geq .16$). Damit lässt sich die deskriptiv leicht positive Veränderung bei den Fünfzehnjährigen und den Neuntklässlerinnen und Neuntklässlern über den beobachteten Zeitraum von sechs Jahren nicht gegen den Zufall absichern.

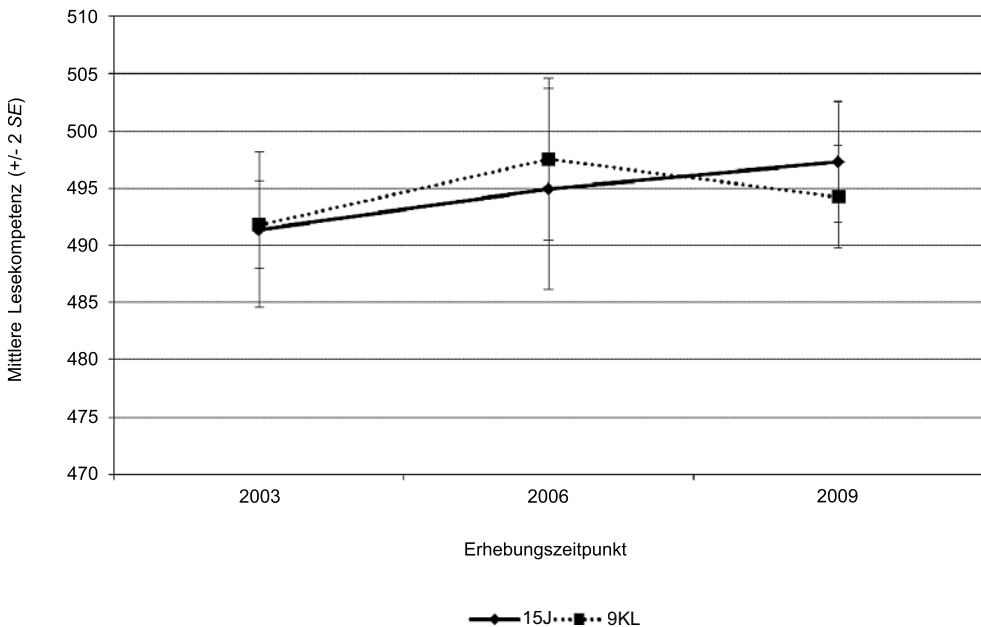


Abb. 1: Mittlere Lesekompetenz in Deutschland von PISA 2003 bis PISA 2009

4 In der Ergebnisdarstellung werden ganzzahlig gerundete Werte angegeben. Es können im Folgenden auf Rundung zurückzuführende vermeintliche Inkonsistenzen bei der Darstellung von Differenzen vorkommen. Alle Berechnungen wurden mit Dezimalstellen durchgeführt, erst für die Ergebnisdarstellung wurde gerundet.

Mathematische Kompetenz

Die Veränderungen der mathematischen Kompetenz in den Stichproben der Fünfzehnjährigen und der neunten Jahrgangsstufe über die Erhebungszeitpunkte sind in Abbildung 2 dargestellt. Die mittlere mathematische Kompetenz der Fünfzehnjährigen in Deutschland lag bei PISA 2003 bei 503 Punkten ($SE = 3.3$ Punkte), bei PISA 2006 bei 504 Punkten ($SE = 3.9$ Punkte) und bei PISA 2009 bei 513 Punkten ($SE = 2.9$ Punkte). Die mittlere mathematische Kompetenz der Fünfzehnjährigen veränderte sich zwischen PISA 2003 und PISA 2006 also kaum, stieg jedoch in den drei darauf folgenden Jahren deutlicher an (siehe auch Frey, Heinze, Mildner, Hochweber & Asseburg, 2010).

In der Stichprobe der neunten Jahrgangsstufe lag die mittlere mathematische Kompetenz bei PISA 2003 bei 504 Punkten ($SE = 1.9$ Punkte). Bei PISA 2006 wurden 501 Punkte ($SE = 3.2$ Punkte) erzielt, bei PISA 2009 507 Punkte ($SE = 2.4$ Punkte). Wie bei den Fünfzehnjährigen erreichte die mittlere mathematische Kompetenz der Neuntklässlerinnen und Neuntklässler deskriptiv ihre höchste Ausprägung bei PISA 2009, bewegte sich aber über alle Erhebungszeitpunkte auf einem sehr ähnlichen Niveau.

Die größten Differenzen finden sich bei den Fünfzehnjährigen mit zehn Punkten zwischen PISA 2009 und PISA 2003 und für die neunte Jahrgangsstufe mit sechs Punkten zwischen PISA 2009 und PISA 2006. Der Anstieg der mathematischen Kompetenz der Fünfzehnjährigen um zehn Punkte zwischen PISA 2003 und PISA 2009 ist statistisch bedeutsam ($p = .04$). Die Effektstärke dieser Differenz beträgt Cohens $d = 0.10$, was nach gängigen Kriterien der psychologischen Forschung als eher geringer Unterschied anzusehen ist (vgl. Bortz & Döring, 2006, S. 606), aber gemessen an den Befun-

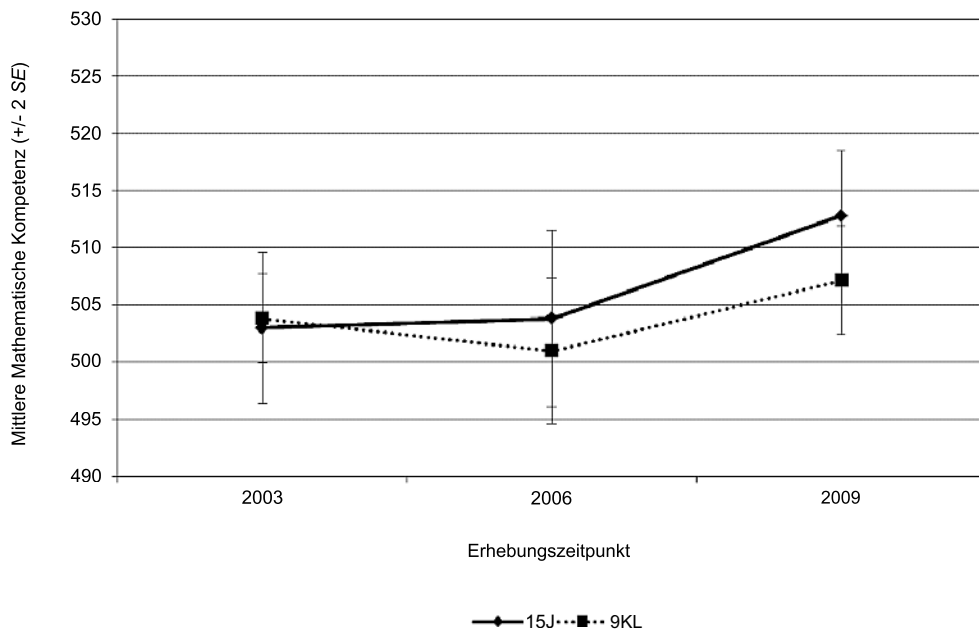


Abb. 2: Mittlere Mathematische Kompetenz in Deutschland von PISA 2003 bis PISA 2009

den von PISA-I-Plus immerhin etwa einem Drittel des Lernzuwachses in der neunten Jahrgangsstufe im Bereich Mathematik entspricht (vgl. Abschnitt 3). Die einzelnen Veränderungen von PISA 2003 zu PISA 2006 (+1 Punkt) beziehungsweise von PISA 2006 zu PISA 2009 (+9 Punkte) sind hingegen nicht signifikant ($p \geq .07$). Für die Stichprobe der neunten Jahrgangsstufe finden sich keinerlei überzufällige Veränderungen zwischen den Erhebungszeitpunkten ($p \geq .14$).

Naturwissenschaftliche Kompetenz

Abbildung 3 zeigt die Veränderungen der naturwissenschaftlichen Kompetenz in den Stichproben der Fünfzehnjährigen und der neunten Jahrgangsstufe zwischen PISA 2006 und PISA 2009. Die mittlere naturwissenschaftliche Kompetenz der Fünfzehnjährigen in Deutschland lag bei PISA 2006 bei 516 Punkten ($SE = 3.8$ Punkte) und bei PISA 2009 bei 520 Punkten ($SE = 2.8$ Punkte). Die mittlere naturwissenschaftliche Kompetenz der Fünfzehnjährigen stieg demnach deskriptiv leicht an (siehe auch Rönnebeck, Schöps, Prenzel, Mildner & Hochweber, 2010).

In der Stichprobe der neunten Jahrgangsstufe lag die mittlere naturwissenschaftliche Kompetenz bei PISA 2006 bei 514 Punkten ($SE = 3.4$ Punkte), bei PISA 2009 wurden 518 Punkte ($SE = 2.5$ Punkte) erreicht. Die mittlere naturwissenschaftliche Kompetenz nahm damit im beobachteten Dreijahreszeitraum auch in der neunten Jahrgangsstufe deskriptiv leicht zu.

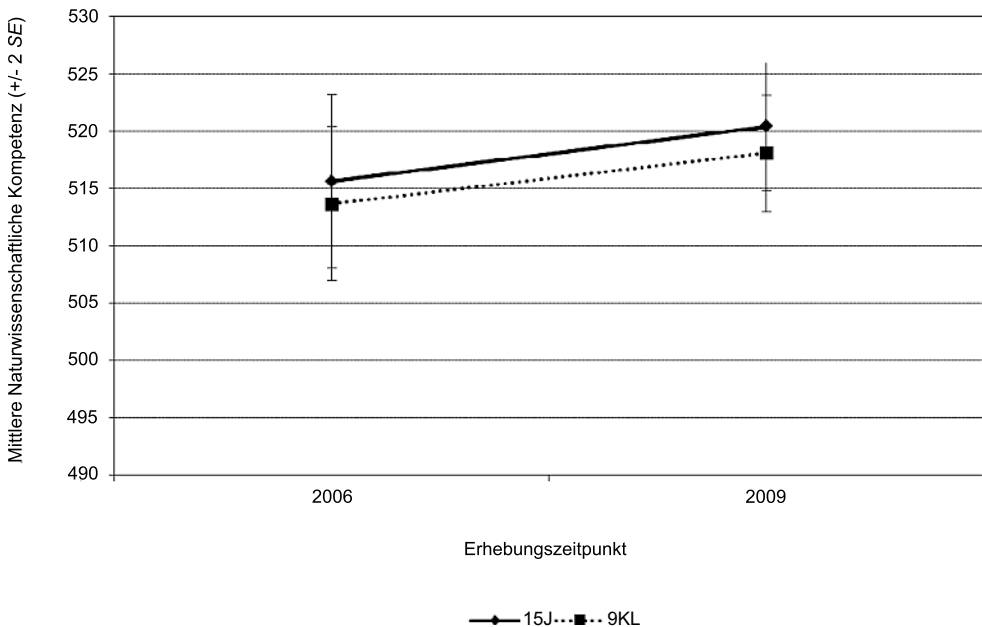


Abb. 3: Mittlere Naturwissenschaftliche Kompetenz in Deutschland von PISA 2006 bis PISA 2009

Die Verläufe sind für die beiden Stichproben nahezu parallel und auch das erreichte Punktniveau ist jeweils sehr ähnlich (vgl. Abbildung 3). Zwischen PISA 2006 und PISA 2009 finden sich deskriptiv sowohl für die Fünfzehnjährigen als auch für die neunte Jahrgangsstufe Zuwächse in Höhe von vier Punkten in der naturwissenschaftlichen Kompetenz. Diese Zuwächse sind vergleichsweise gering und statistisch nicht signifikant ($p \geq .37$).

Zusammenfassung

Deskriptiv betrachtet erweisen sich die Verläufe der mittleren Kompetenzen über die beobachteten Erhebungszeitpunkte in den Stichproben der Fünfzehnjährigen und der Neuntklässlerinnen und Neuntklässler als relativ ähnlich, wobei die Veränderung der mittleren Kompetenzen bei den Fünfzehnjährigen insgesamt etwas günstiger ausfällt. Allerdings lässt sich in beiden Stichproben keine der Veränderungen zwischen PISA 2006 und PISA 2009 gegen den Zufall absichern; ebenso statistisch nicht signifikant sind die Veränderungen zwischen PISA 2003 und PISA 2006, die auch deskriptiv in beiden Stichproben gering ausfallen. Die einzige signifikante Veränderung zeigt sich zwischen PISA 2003 und PISA 2009 in der mittleren mathematischen Kompetenz der Fünfzehnjährigen – ein positiver Trend, der sich für die Neuntklässlerinnen und Neuntklässler nicht nachweisen lässt.

6.2 Analysen zur Verteilung der Fünfzehnjährigen auf die Jahrgangsstufen

Können die beobachteten Trends in den Kompetenzmittelwerten der Fünfzehnjährigen mit den Veränderungen in der Verteilung auf die Jahrgangsstufen in Beziehung gebracht werden? Es zeigt sich, dass die in Abschnitt 4 formulierten Annahmen im Wesentlichen zutreffen: Zwischen den PISA-Erhebungen 2003 und 2009 beziehungsweise 2006 und 2009 zeigen sich im Lesen und in den Naturwissenschaften deskriptiv nur geringe positive Veränderungen, die sich nicht statistisch absichern lassen. In der Mathematik resultiert dagegen zwischen PISA 2003 und PISA 2009 ein deutlicher sowie signifikanter mittlerer Kompetenzzuwachs. Der positive Trend in der Mathematik korrespondiert mit einer Verringerung der Anteile von Achtklässlerinnen und Achtklässlern sowie Neuntklässlerinnen und Neuntklässlern und einer Zunahme des Anteils der Zehntklässlerinnen und Zehntklässler zwischen PISA 2003 und PISA 2009. Berücksichtigt man Studien, die bereichs- sowie jahrgangsstufenspezifische Lernzuwächse von Schülerinnen und Schülern ermitteln konnten (vgl. Abschnitt 3), erscheint es plausibel, dass sich dies gerade im Bereich der Mathematik leistungsförderlich ausgewirkt hat.

Um einen noch genaueren Eindruck zu erhalten, wurde eine vertiefende deskriptive Analyse basierend auf neu erzeugten Gewichten der Schülerinnen und Schüler durchgeführt (vgl. Abschnitt 5.2). Dabei wurde als wichtige Kontrollvariable die Schulartzugehörigkeit berücksichtigt. Differenziert wurde hier zwischen dem Gymnasium einerseits und den nicht-gymnasialen Schularten andererseits, da das Gymnasium als einzige

Schulart im interessierenden Zeitraum und in allen Bundesländern von bildungsstrukturellen Reformmaßnahmen relativ unberührt geblieben ist. Die neuen Gewichte der Schülerinnen und Schüler wurden gebildet, indem die originalen Gewichte der PISA-Erhebungen 2006 und 2009 (a) je Schulart (Gymnasium versus nicht-gymnasiale Schularten) und (b) je Schulart und Jahrgangsstufe mit dem Quotienten der Gewichtssumme in PISA 2003 und der Gewichtssumme in PISA 2006 beziehungsweise PISA 2009 multipliziert wurden. Basierend auf den neu generierten Gewichten wurden dann die Kompetenzmittelwerte in den drei Domänen für die PISA-Erhebungen 2006 und 2009 neu berechnet. Die Resultate lassen sich als (fiktive) Ergebnisse bei Konstanthalten (a) der Schulart beziehungsweise (b) der Schulart und der Jahrgangsstufenverteilung für die Fünfzehnjährigen seit PISA 2003 interpretieren. Die Ergebnisse mit den originalen und den neu erzeugten Gewichten der Schülerinnen und Schüler sind in Tabelle 3 gegenübergestellt.

Domäne/Analyse	PISA 2003	PISA 2006	PISA 2009
Lesen			
Mit Originalgewichten	491	495	497
Neu gewichtet: Schulart	–	493	492
Neu gewichtet: Schulart + Jahrgangsstufe	–	489	489
Mathematik			
Mit Originalgewichten	503	504	513
Neu gewichtet: Schulart	–	502	507
Neu gewichtet: Schulart + Jahrgangsstufe	–	498	502
Naturwissenschaften			
Mit Originalgewichten	–	516	520
Neu gewichtet: Schulart	–	513	515
Neu gewichtet: Schulart + Jahrgangsstufe	–	511	511

Tab. 3: *Mittlere Kompetenzen der Fünfzehnjährigen in den PISA-Erhebungen 2003 bis 2009 bei Verwendung der originalen Stichprobengewichte und bei Neugewichtung der PISA-Erhebungen 2006 und 2009 basierend auf (a) der Schulart sowie (b) der Schulart und der Jahrgangsstufenverteilung der Fünfzehnjährigen in PISA 2003*

Der Vergleich der Befunde mit den originalen und den neuen Gewichten weist darauf hin, dass der signifikante positive Trend bei den Fünfzehnjährigen in der Mathematik zwischen PISA 2003 und PISA 2009 bereits in erheblichem Umfang durch Veränderungen in der Schulartverteilung erklärt werden kann – die Differenz zwischen PISA 2003 und PISA 2009 beträgt bei Berücksichtigung der Schulart nur noch vier Punkte. Da-

rüber hinaus erweisen sich aber auch die Veränderungen in der Jahrgangsstufenverteilung als bedeutsam, es ergibt sich bei Einbezug beider Merkmale eine annähernde Nivellierung der beobachteten Punktedifferenz. Auch für die mittleren Lesekompetenzen sowie die mittleren naturwissenschaftlichen Kompetenzen, bei denen allerdings keine statistische Absicherung des Trends möglich war (vgl. Abschnitt 6.1), resultiert bei Berücksichtigung der Schulart eine deutliche Verringerung der Punktedifferenz zwischen den PISA-Erhebungen. In den Naturwissenschaften, nicht aber im Bereich des Lesens, führt die zusätzliche Berücksichtigung der Jahrgangsstufenverteilung bei der Bildung neuer Stichprobengewichte zu einer weitgehenden Angleichung der ermittelten Kompetenzstände für die Erhebungszeitpunkte.

7. Fazit und Ausblick

In dem vorliegenden Beitrag wurden Trends der mittleren Kompetenzen in den Domänen Lesen, Mathematik und Naturwissenschaften in den PISA-Stichproben der Fünfzehnjährigen und der Neuntklässlerinnen und Neuntklässler vergleichend analysiert. Herangezogen wurden dazu die Ergebnisse der PISA-Erhebungen der Jahre 2003, 2006 und 2009 (Lesen und Mathematik) beziehungsweise der Jahre 2006 und 2009 (Naturwissenschaften). Als ein schulstrukturelles Merkmal, das zur Erklärung der Trends bei den Fünfzehnjährigen und von möglichen differentiellen Trends bei Fünfzehnjährigen und Neuntklässlerinnen und Neuntklässlern geeignet erscheint, wurde exemplarisch die Verteilung der Fünfzehnjährigen auf die Jahrgangsstufen untersucht.

Für die Stichproben der Fünfzehnjährigen und der Neuntklässlerinnen und Neuntklässler lässt sich im Lesen und in den Naturwissenschaften keine der Differenzen zwischen den Erhebungszeitpunkten zufallskritisch absichern. Dagegen zeigt sich für die Fünfzehnjährigen, anders als für die Neuntklässlerinnen und Neuntklässler, in der Mathematik ein signifikanter positiver Trend zwischen PISA 2003 und PISA 2009. Diese Disparität weist darauf hin, dass die mit den deutschen PISA-Daten möglichen alters- und jahrgangsstufenbasierten Trendanalysen durchaus zu unterschiedlichen Befunden führen können. Entsprechend könnte es interessant sein, die Trendberichterstattung bei PISA in Deutschland um Ergebnisse für die Neuntklässlerinnen und Neuntklässler zu ergänzen. Die dadurch verfügbaren zusätzlichen Informationen besitzen praktische Relevanz. So liegt beispielsweise in der Unterrichtsforschung – für welche die Organisation von Schule in Jahrgangsstufen eine zentrale Rolle spielt – die Bezugnahme auf die Kompetenzen der Neuntklässlerinnen und Neuntklässler näher als auf jene der Fünfzehnjährigen.

In Übereinstimmung mit den von uns formulierten Erwartungen zeigte sich, dass die domänenspezifischen Trends bei den Fünfzehnjährigen zumindest deskriptiv mit Veränderungen in der Verteilung auf die Jahrgangsstufen in Verbindung zu bringen sind. Die Tendenz hin zum Besuch höherer Jahrgangsstufen korrespondierte erwartungsgemäß insbesondere mit einem bedeutsamen positiven Trend im Bereich Mathematik, während im Lesen kein vergleichbarer Kompetenzzuwachs zu verzeichnen war. Für die Natur-

wissenschaften konnten keine eindeutigen Erwartungen abgeleitet werden. Die Analyse beschränkte sich hier – in Übereinstimmung mit dem Vorgehen der OECD bei der internationalen Trendberichterstattung (OECD, 2010) – auf die PISA-Erhebungen 2006 und 2009. Für diesen Zeitraum resultierte ein deskriptiv schwacher und nicht signifikanter positiver Trend, der mit eher geringfügigen Veränderungen in der Jahrgangsstufenverteilung korrespondierte. Die potenzielle Relevanz der Jahrgangsstufenverteilung für die Erklärung der Trends bei den Fünfzehnjährigen konnte in den Analysen unter Verwendung neuer Stichprobengewichte auch unter Hinzunahme der Schulart (Gymnasium versus nicht-gymnasiale Schularten) gestützt werden. Für alle Domänen resultierte, dass die beobachteten Punktedifferenzen weitgehend verschwinden, wenn Veränderungen in der Schulart- sowie Jahrgangsstufenverteilung der Fünfzehnjährigen berücksichtigt werden. Dies könnte darauf hindeuten, dass sich die bildungsstrukturellen Maßnahmen, die mehr Fünfzehnjährigen den Besuch höherer Jahrgangsstufen ermöglichen, tatsächlich in positiven Veränderungen in den Kompetenzmittelwerten (zumindest in der Mathematik) niedergeschlagen haben. Die Kompetenzen der Neuntklässlerinnen und Neuntklässler erwiesen sich demgegenüber als nahezu unverändert, die Differenzen zwischen den Erhebungszeitpunkten waren zudem statistisch nicht abzusichern. Allerdings ist bei solchen Schlussfolgerungen große Vorsicht geboten, da die beobachteten Beziehungen weder statistisch abgesichert sind, noch weitere möglicherweise einflussreiche Faktoren berücksichtigt wurden.

Zusätzliche potenzielle Einschränkungen bei der Interpretation der Ergebnisse dieser Studie ergeben sich aus den bereits angesprochenen Befunden von Carstensen et al. (2008), die zeigen konnten, dass alternative Skalierungsverfahren zu voneinander abweichenden Trendergebnissen für die Lesekompetenz der Fünfzehnjährigen in Deutschland führen (vgl. Abschnitt 5.2). Demnach wurden in den von der OECD berichteten Trends für die PISA-Erhebungen 2000 bis 2006 die Kompetenzen der Schülerinnen und Schüler für PISA 2003 und PISA 2006 überschätzt. Welche Implikationen diese Befunde für den vorliegenden Beitrag besitzen, lässt sich ohne die Durchführung vergleichbarer Analysen unter Einschluss von PISA 2009 nicht abschließend klären. Allerdings könnten auch im vorliegenden Fall abweichende Befunde resultieren, die sich – angesichts der Verwendung der internationalen Aufgabenschwierigkeitsparameter in den entsprechenden Skalierungen – auch auf die Stichprobe der Neuntklässlerinnen und Neuntklässler erstrecken könnten.

Die von uns durchgeführte Studie könnte in mindestens zwei Richtungen erweitert werden. Erstens haben wir uns in unseren Analysen auf die Betrachtung von Trends in den Kompetenzmittelwerten der Schülerinnen und Schüler beschränkt. Neben den Kompetenzmittelwerten gibt es noch andere statistische Kennwerte, die für die Beurteilung von Kompetenzverläufen bedeutsam sind und in zukünftigen Untersuchungen einbezogen werden könnten, insbesondere die Streuung der Kompetenzen. Zweitens haben wir als einen möglichen Faktor zur Erklärung der Veränderungen in den Kompetenzen bei den Fünfzehnjährigen und von differentiellen Trends bei Fünfzehnjährigen und Neuntklässlerinnen und Neuntklässlern die Verteilung der Fünfzehnjährigen auf die Jahrgangsstufen betrachtet. Unsere Analysen konnten dabei erste Anhaltspunkte

für mögliche Erklärungszusammenhänge liefern, blieben aber letztlich rein deskriptiv und erlaubten es nicht, die Beziehung zwischen Trendverläufen und potenziellen Bedingungsmerkmalen inferenzstatistisch abzusichern. Eine noch aussagekräftigere Auswertung wäre beispielsweise auf Grundlage einer gemeinsamen (marginalen) Skalierung (siehe Carstensen et al., 2008) der verschiedenen PISA-Erhebungen möglich. Dabei könnten die latenten Differenzen zwischen den Erhebungszeitpunkten direkt mit möglichen Erklärungsvariablen in Beziehung gesetzt werden.

Literatur

- Baumert, J., Stanat, P., & Demmrich, A. (2001). PISA 2000: Untersuchungsgegenstand, theoretische Grundlagen und Durchführung der Studie. In J. Baumert, E. Klieme, M. Neubrand, M. Prenzel, U. Schiefele, W. Schneider, P. Stanat, K.-J. Tillmann & M. Weiß (Hrsg.), *PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich* (S. 15-68). Opladen: Leske + Budrich.
- Baumert, J., & Stanat, P. (2006). Internationale Schulleistungsvergleiche. In D. H. Rost (Hrsg.), *Handwörterbuch Pädagogische Psychologie* (S. 291-302). Weinheim/Basel: Beltz.
- Bortz, J., & Döring, N. (2006). *Forschungsmethoden und Evaluation*. Heidelberg: Springer.
- Carstensen, C. H., Frey, A., Walter, O., & Knoll, S. (2007). Technische Grundlagen des dritten internationalen Vergleichs. In M. Prenzel, C. Artelt, J. Baumert, W. Blum, M. Hammann, E. Klieme & R. Pekrun (Hrsg.), *PISA 2006. Die Ergebnisse der dritten internationalen Vergleichsstudie* (S. 367-390). Münster: Waxmann.
- Carstensen, C. H., Prenzel, M., & Baumert, J. (2008). Trendanalysen in PISA: Wie haben sich die Kompetenzen in Deutschland zwischen PISA 2000 und PISA 2006 entwickelt? *Zeitschrift für Erziehungswissenschaft, Sonderheft 10*, 11-34.
- Ehmke, T., Blum, W., Neubrand, M., Jordan, A., & Ulfig, F. (2006). Wie verändert sich die mathematische Kompetenz von der neunten zur zehnten Klassenstufe? In M. Prenzel, J. Baumert, W. Blum, R. Lehmann, D. Leutner, M. Neubrand, R. Pekrun, J. Rost & U. Schiefele (Hrsg.), *PISA 2003. Untersuchungen zur Kompetenzentwicklung im Verlaufe eines Schuljahres* (S. 63-86). Münster: Waxmann.
- Frey, A., Heinze, A., Mildner, D., Hochweber, J., & Asseburg, R. (2010). Mathematische Kompetenz von PISA 2003 bis PISA 2009. In E. Klieme, C. Artelt, J. Hartig, N. Jude, O. Köller, M. Prenzel, W. Schneider & P. Stanat (Hrsg.), *PISA 2009. Bilanz nach einem Jahrzehnt* (S. 153-176). Münster: Waxmann.
- Gailberger, S., & Willenberg, H. (2008). Leseverstehen Deutsch. In DESI-Konsortium (Hrsg.), *Unterricht und Kompetenzerwerb in Deutsch und Englisch. Ergebnisse der DESI-Studie* (S. 60-71). Weinheim/Basel: Beltz.
- Gebhardt, E., & Adams, R. J. (2007). The influence of equating methodology on reported trends in PISA. *Journal of Applied Measurement*, 8(3), 305-322.
- Helmke, A. (2010). *Unterrichtsqualität und Lehrprofessionalität. Diagnose, Evaluation und Verbesserung des Unterrichts*. Seelze: Kallmeyer.
- Herzberg, P. Y., & Frey, A. (2011). Kriteriumsorientierte Diagnostik. In L. Hornke, M. Amelang & M. Kersting (Hrsg.), *Enzyklopädie der Psychologie*, Themenbereich B Methodologie und Methoden, Serie II Psychologische Diagnostik, Band 2 Methoden der psychologischen Diagnostik (S. 281-324). Göttingen: Hogrefe.
- Klieme, E., Jude, N., Baumert, J., & Prenzel, M. (2010). PISA 2000-2009: Bilanz der Veränderungen im Schulsystem. In E. Klieme, C. Artelt, J. Hartig, N. Jude, O. Köller, M. Prenzel, W. Schneider & P. Stanat (Hrsg.), *PISA 2009. Bilanz nach einem Jahrzehnt* (S. 277-296). Münster: Waxmann.

- KMK (Hrsg.) (2004a). *Bildungsstandards im Fach Mathematik für den Mittleren Schulabschluss. Beschluss vom 4.12.2003*. München: Luchterhand.
- KMK (Hrsg.) (2004b). *Bildungsstandards im Fach Deutsch für den Mittleren Schulabschluss. Beschluss vom 4.12.2003*. München: Luchterhand.
- KMK (Hrsg.) (2004c). *Bildungsstandards für die erste Fremdsprache (Englisch/Französisch) für den Mittleren Schulabschluss. Beschluss vom 4.12.2003*. München: Luchterhand.
- Köller, O., Knigge, M., & Tesch, B. (Hrsg.) (2010). *Sprachliche Kompetenzen im Ländervergleich: Befunde des ersten Ländervergleichs zur Überprüfung der Bildungsstandards für den Mittleren Schulabschluss in den Fächern Deutsch, Englisch und Französisch*. Münster: Waxmann.
- Lehmann, R., Peek, R., Gänsfuß, R., & Husfeldt, V. (2002). *LAU 9. Aspekte der Lernausgangslage und der Lernentwicklung – Klassenstufe 9. Ergebnisse einer Längsschnittuntersuchung in Hamburg*. Hamburg: Behörde für Bildung und Sport.
- Naumann, J., Artelt, C., Schneider, W., & Stanat, P. (2010). Lesekompetenz von PISA 2000 bis PISA 2009. In E. Klieme, C. Artelt, J. Hartig, N. Jude, O. Köller, M. Prenzel, W. Schneider & P. Stanat (Hrsg.), *PISA 2009. Bilanz nach einem Jahrzehnt* (S. 23-72). Münster: Waxmann.
- OECD (2006). *Assessing scientific, reading and mathematical literacy: A framework for PISA 2006*. Paris: OECD.
- OECD (2010). *PISA 2009 results: Learning trends: Changes in student performance since 2000 (Volume V)*. Paris: OECD.
- OECD (2011). *PISA 2009 technical report*. Paris: OECD.
- Prenzel, M., Baumert, J., Blum, W., Lehmann, R., Leutner, D., Neubrand, M., Pekrun, R., Rost, J., & Schiefele, U. (Hrsg.) (2006). *PISA 2003. Untersuchungen zur Kompetenzentwicklung im Verlauf eines Schuljahres*. Münster: Waxmann.
- Rönnebeck, S., Schöps, K., Prenzel, M., Mildner, D., & Hochweber, J. (2010). Naturwissenschaftliche Kompetenz von PISA 2006 bis PISA 2009. In E. Klieme, C. Artelt, J. Hartig, N. Jude, O. Köller, M. Prenzel, W. Schneider & P. Stanat (Hrsg.), *PISA 2009. Bilanz nach einem Jahrzehnt* (S. 177-198). Münster: Waxmann.
- Seidel, T., & Shavelson, R. J. (2007). Teaching effectiveness research in the last decade: Role of theory and research design in disentangling meta-analysis results. *Review of Educational Research*, 77(4), 454-499.
- Sibberns, H., & Baumert, J. (2001). Anhang A: Stichprobenziehung und Stichprobengewichtung. In J. Baumert, E. Klieme, M. Neubrand, M. Prenzel, U. Schiefele, W. Schneider, P. Stanat, K.-J. Tillmann & M. Weiß (Hrsg.), *PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich* (S. 511-524). Opladen: Leske + Budrich.
- von Davier, M., Gonzalez, E., & Mislevy, R. J. (2009). What are plausible values and why are they useful? *IERI Monograph Series*, 4, 9-36.

Anschrift der Autorin/der Autoren

Dipl.-Psych. Dorothea Mildner, Deutsches Institut für Internationale Pädagogische Forschung (DIPF), Schloßstraße 29, 60486 Frankfurt am Main, Deutschland
E-Mail: mildner@dipf.de

Dr. Jan Hochweber, Deutsches Institut für Internationale Pädagogische Forschung (DIPF), Schloßstraße 29, 60486 Frankfurt am Main, Deutschland
E-Mail: hochweber@dipf.de

Prof. Dr. Andreas Frey, Friedrich-Schiller-Universität Jena, Institut für Erziehungswissenschaft, Am Planetarium 4, 07737 Jena, Deutschland
E-Mail: andreas.frey@uni-jena.de

Schulentwicklung im Verlauf eines Jahrzehnts – Erste Ergebnisse des PISA-Schulpanels

Das „Programme for International Student Assessment“ (PISA) ist ein internationales Projekt, in dem alle drei Jahre die Kompetenzen von fünfzehnjährigen Schülerinnen und Schülern erfasst werden. Die Erfassung der Kompetenzen erfolgt in jedem Durchgang mit anderem Schwerpunkt; je drei Erhebungen mit Schwerpunkten in Lesen, Mathematik und Naturwissenschaften bilden einen Zyklus. Im Jahr 2009 stand mit Beginn des zweiten Zyklus erstmals seit PISA 2000 wieder die Messung der Lesekompetenz (*Reading Literacy*) der Schülerinnen und Schüler im Mittelpunkt. Während die PISA-Ergebnisse Aufschluss über Veränderungen in den Kompetenzen sowie bestimmten weiteren Merkmalen (z.B. Lernmotivation) fünfzehnjähriger Schülerinnen und Schüler in den beteiligten Staaten geben können, im Sinne von Unterschieden zu vorherigen Kohorten, informieren sie weder über die Entwicklung einzelner Schülerinnen und Schüler noch über die Entwicklung von Schulen. Dies liegt im Design der PISA-Studie begründet. Schulen werden zu jedem Messzeitpunkt unabhängig voneinander zufällig gezogen und gelangen entsprechend nur zufällig mehrmals in die PISA-Stichprobe (Klieme & Steiner, 2008). PISA erlaubt also keine echten längsschnittlichen Aussagen über die Entwicklung von Schülerinnen und Schülern bzw. Schulen, sondern nur Trendaussagen auf der Ebene der Staaten. So können die Frage nach den Bedingungsfaktoren für individuelle Leistungsentwicklung und die Frage nach Bedingungsfaktoren für die Entwicklung des Leistungsniveaus auf Schulebene nicht befriedigend beantwortet werden. Hierzu sind Erweiterungen des PISA-Designs erforderlich, die bislang nur innerhalb Deutschlands realisiert werden konnten (vgl. Klieme, Jude, Baumert & Prenzel, 2010, S. 295).

Eine Möglichkeit, über den Zeitraum seit der ersten PISA-Erhebung hinweg längsschnittliche Informationen zu Entwicklungen auf Schulebene zu erhalten, ergab sich mit dem PISA-Schulpanel. Ergänzend zur internationalen Stichprobe für PISA 2009 wurde in Deutschland national eine Zusatzstichprobe von 59 Gymnasien gezogen, die auch an PISA 2000 teilgenommen hatten.¹ Die Untersuchung beschränkte sich auf Gymnasien, da das Gymnasium eine Schulart ist, die in jedem Bundesland in Deutschland existiert, und weil Umwandlungen der Schulart bei Gymnasien in diesem Zeitraum so gut wie nie vorkamen. Die Schulpanel-Stichprobe ermöglicht die längsschnittliche Analyse der Entwicklung der teilnehmenden Gymnasien über einen Zeitraum von neun Jahren. Im vorliegenden Beitrag wurde untersucht, in welchem Zusammenhang ausgewählte

1 Da wir uns hier nur mit deutschen Daten beschäftigen, unterscheiden wir nicht zwischen den Schulen, die bei PISA 2000 in den internationalen Vergleich eingingen (Baumert et al., 2001), und der erweiterten Stichprobe PISA-E, die Grundlage des Vergleichs zwischen den Bundesländern war (Baumert et al., 2003).

Schulmerkmale und deren Veränderung mit der Veränderung kognitiver und nicht-kognitiver Merkmale von Schülerinnen und Schülern dieser Schulen stehen.

Im Folgenden gehen wir nach der Darstellung des Hintergrunds unseres Beitrags (Abschnitt 1) auf die Zielsetzung und das Design der Studie (Abschnitt 2) sowie die konkreten Fragestellungen dieses Artikels ein (Abschnitt 3) und beschreiben anschließend unser methodisches Vorgehen (Abschnitt 4). In den Abschnitten 5 und 6 werden die Ergebnisse dargestellt und diskutiert.

1. Theoretischer Hintergrund: Von der Schuleffektivitäts- zur Schulentwicklungsforschung

Die vorliegende Untersuchung befasst sich mit der Entwicklung der Schule als Organisation mit dem Ziel, Prädiktoren der Entwicklung von Schulen zu identifizieren. Diese Fragestellung ist der herkömmlichen Schuleffektivitäts- bzw. Schulqualitätsforschung verwandt, beinhaltet aber einen fundamentalen Wechsel der Perspektive. Daher unterscheiden wir im Folgenden explizit zwischen Aspekten der Schuleffektivität (Abschnitt 1.1) und der Schulqualität (Abschnitt 1.2) und schließen dann den Bogen zur Schulentwicklungsforschung (Abschnitt 1.3), um anschließend die hier untersuchten Prädiktoren der Entwicklung von Schulen näher zu erläutern (Abschnitt 1.4).

1.1 Schuleffektivität

Die Schuleffektivität wird üblicherweise daran gemessen, welche Ergebnisse eine Schule erreicht („Output“), wobei meist in Rechnung gestellt wird, unter welchen Bedingungen eine Schule arbeitet („Kontext“ und „Input“). Die Auswahl der Kriterien zur Messung effektiver Schulen, also der untersuchten Outputmerkmale, wird immer wieder diskutiert (de Maeyer, van den Bergh, Rymenans, van Petegem & Rijlaarsdam, 2010; Sammons, Hillman & Mortimore, 1995; Scheerens & Bosker, 1997; Seidel, 2008; Teddlie & Reynolds, 2000). Die Auswahl eines einzelnen Effektivitätskriteriums ergibt kein generalisierbares Bild davon, welche Merkmale der Schule eine effektivitätssteigernde Wirkung haben. Dies bedarf einer breiteren Auswahl an Kriterien.

Kriterien effektiver Schulen lassen sich unterschiedlich einteilen. Seidel (2008) beschreibt vier Kategorien von Effektivitätskriterien: kognitive, motivational-affektive, metakognitive und verhaltensbezogene Kriterien. Eine einfachere Einteilung ist die in kognitive und non-kognitive Effektivitätskriterien (Mortimore, Sammons & Thomas, 1994; Opdenakker & van Damme, 2000; van Damme et al., 2006). Zumeist wird die Effektivität von Schulen ausschließlich über kognitive, d.h. über leistungs- bzw. kompetenzbezogene Kriterien gemessen (Scheerens, 2008; Scheerens & Bosker, 1997; Seidel, 2008; van Damme et al., 2006), wobei sehr unterschiedliche Leistungsmaße als Kriterien herangezogen werden, z.B. die Leistung in bestimmten Fächern oder Kompetenzen (vgl. Seidel, 2008). Im Kontext des PISA-Schulpanels stützen wir uns als kognitives

Kriterium auf die Lesekompetenz der Schülerinnen und Schüler, da diese in 2000 und 2009 als Schwerpunkt erhoben wurde.

Die neuere Schulforschung berücksichtigt zunehmend auch non-kognitive, affektive Kriterien (vgl. Diedrich, 2008; Gerecht, 2010; Opdenakker & van Damme, 2000). Kognitive und non-kognitive Kriterien können durchaus zu unterschiedlichen Schlussfolgerungen führen. So wird bei van Damme et al. (2006) deutlich gemacht, dass Schulen einen Einfluss auf non-kognitive Kriterien haben; und dass Schulen, die hinsichtlich eines kognitiven Kriteriums als effektiv eingestuft werden, nicht notwendigerweise auch einen positiven Effekt auf non-kognitive Kriterien wie beispielsweise das Wohlbefinden haben müssen. Im PISA-Schulpanel verwenden wir als ein non-kognitives Kriterium die von den Schülerinnen und Schülern wahrgenommene Beziehungsqualität zu ihren Lehrerinnen und Lehrern (kurz: Beziehungsqualität). Diese kann in Anlehnung an van Damme, de Fraine, van Landeghem, Opdenakker und Onghena (2002) als ein Aspekt des Effektivitätskriteriums Schülerwohlbefinden aufgefasst werden. Diese Einordnung der Beziehungsqualität als Indikator für Schülerwohlbefinden kann kritisch betrachtet werden, denn die Beziehungsqualität wurde, beispielsweise in PISA 2009, als Indikator für das Schulklima herangezogen (vgl. Hertel, Hochweber, Steinert & Klieme, 2010; siehe auch Eder, 2002). Das Klima einer Schule ist allerdings streng genommen kein Effektivitätskriterium wie das Leistungsniveau der Schülerinnen und Schüler, sondern ein Prozessmerkmal der Schule und somit ein Merkmal der Schulqualität. Verschiedene Forscher fanden jedoch einen positiven Effekt eines guten Schulklimas auf Schülermerkmale, wie beispielsweise die Lernmotivation der Schülerinnen und Schüler (Brophy, 2000; Fraser, 1994; Götz, Frenzel & Pekrun, 2008). Zudem ist das Schulklima „sensibel für Veränderungen im Schulsystem bzw. Entwicklungsmaßnahmen einzelner Schulen“ (Eder, 2006, S. 626). Dies bestärkt uns darin, neben Veränderungen im Leseleistungsniveau Veränderungen in der Beziehungsqualität als Kriterium für die Entwicklung der Schulen zu nutzen.

1.2 Schulqualität

Schulqualität wird in der Forschungsliteratur manchmal gleichgesetzt mit Schuleffektivität. Präziser ist jedoch eine klare Trennung: Als Aspekte der Schulqualität bezeichnen wir jene Merkmale der pädagogischen Arbeit und der Schulorganisation, also „Prozessmerkmale“, die Schuleffektivität (direkt oder indirekt) im Sinne eines hohen Outputs ermöglichen. Die Schulforschung hat vielfältige Faktoren identifiziert, die in diesem Sinne „gute Schulen“ ausmachen. Eine Liste solcher Faktoren wurde auf der Grundlage praktischer Erfahrungen erstmals von Edmonds (1979) vorgeschlagen: Qualität des Curriculums – damals gleichgesetzt mit einer Betonung von „Basic Skills“ –, hohe Leistungserwartungen, Schulklima, Führungsqualität der Schulleitung sowie die Nutzung von Leistungsmessungen und Evaluation. Neuere Übersichten, die sich auf Meta-Analysen zahlreicher empirischer Studien stützen, bestätigen diese Faktoren und ergänzen sie um Konsens und Kooperation der Lehrerinnen und Lehrer sowie die Einbindung

der Eltern (so etwa Creemers & Reezigt, 1996; Scheerens, 2008). Scheerens (2008) führt zudem effektivitätssteigernde Faktoren an, die sich auf den Unterricht beziehen: Klassenklima, Klassenmanagement/effektive Lernzeit, strukturierte Instruktionen, Differenzierung und adaptive Instruktionen sowie Feedback. Creemers und Reezigt (1996) hingegen benennen ergänzende Faktoren der Schulebene, die guten Unterricht ermöglichen: Fokus auf Grundfertigkeiten/Lernzeit sowie die Koordination von Curriculum und Instruktion über Klassen und Jahrgangsstufen hinweg.

Ausgehend vor allem von den Arbeiten Fends (z.B. 1998) liegen inzwischen auch im deutschen Sprachraum verschiedene Modelle vor, die diesen Ansatz weiter ausdifferenzieren. Ditton (2000a, 2000b, 2007) unterscheidet Individual-, Unterrichts-, Schul- und Kontextebene und bezeichnet sein Modell als „4-plus-Modell“, da die Kontextebene noch genauer unterteilt werden kann. Den auf den verschiedenen Ebenen verorteten Faktoren wurde in der Schulqualitätsforschung eine unterschiedliche Wirksamkeit insbesondere bezüglich der Leistungsergebnisse der Schülerinnen und Schüler zugesprochen. Ditton (2000a, 2000b) verweist hier auf die „These der primären Bedeutung proximaler Faktoren“ (d.h. den Schülerinnen und Schülern näher stehender Faktoren), nach welcher die direkte Lernumgebung der Schülerinnen und Schüler den größten Effekt auf die Schülerleistung hat. Diese These wird beispielsweise von Klieme, Steinert und Hochweber (2010) gestützt, die nachweisen, dass Kooperation und geteilte Ziele in Deutsch-Fachkollegien zwar den Unterricht vielfältiger machen und Disziplinprobleme verringern, aber nicht auf die individuelle Leistungs- und Motivationsentwicklung der Schülerinnen und Schüler durchschlagen.

Gemeinhin werden Faktoren, die in Schulqualitätsstudien als wirksam identifiziert wurden, auch als bedeutsam für die Schulentwicklung angesehen. Im vorliegenden Beitrag wollen wir unser Augenmerk auf proximale und distale Faktoren richten, die in den letzten Jahren auch bildungspolitisch eine hohe Aufmerksamkeit bekommen haben und für die im Schulpanel Veränderungsmaße vorliegen, um zu prüfen, welche Rolle sie für die Entwicklung von Schulen spielen. Konkret werden in unserer Untersuchung individuelle Förderung (als proximaler Faktor für Lernprozesse) sowie ganztägige Schulorganisation und Evaluationsaktivitäten (als distale Faktoren) als Prädiktoren einbezogen.

1.3 Schulentwicklung

Anders als die Schuleffektivitäts- und Schulqualitätsforschung, für welche die Wirkung von Schule (für den einzelnen Schüler/die einzelne Schülerin) in einer bestimmten Kohorte, d.h. zu einem gegebenen Zeitpunkt, im Mittelpunkt steht, zielt die Schulentwicklungsforschung auf die Beschreibung und Prädiktion der Veränderungen von Schulen ab. Die Bedeutung dieses Perspektivwechsels ist in der Schulforschung allerdings bislang kaum systematisch aufgearbeitet (vgl. dazu Klieme & Steinert, 2008). Selbst neuere Ansätze zu einem „Dynamic model of school effectiveness“ (Creemers & Kyriakides, 2008) oder einer Theorie der *Schulgestaltung* (Fend, 2008) stützen sich

nach wie vor stark auf die querschnittliche Schulforschung (wozu wir hier auch Studien zählen, die *Schülerinnen und Schüler* wiederholt testen, sofern sie die Schul- und Unterrichtsmerkmale nur einmal messen) und auf Analysen effektiver Schulen, während längsschnittliche Erhebungen nur in wenigen Fällen vorgenommen wurden. In der Methodologie sozialwissenschaftlicher Forschung ist hinlänglich bekannt, dass Entwicklungsfragen nicht mit Querschnittsdaten untersucht werden können (Steinhage & Blossfeld, 1999). Längsschnittliche Betrachtungen von Schulen sind beispielsweise unerlässlich, um rekursive Effekte, d.h. wechselseitige Abhängigkeiten zwischen Bedingungsfaktoren und Kriterien, identifizieren zu können. Beispiele erster deutscher Untersuchungen mit längsschnittlichem Design auf Schulebene sind die Studie zur Entwicklung von Ganztagschulen (StEG; vgl. Klieme, Fischer, Holtappels, Rauschenbach & Stecher, 2010), die Begleitforschung zum Projekt „Selbstständige Schule“ (Holtappels, Klemm & Rolff, 2008) sowie die Untersuchung von Trautwein, Neumann, Nagy, Lüdtke und Maaz (2010) zur gymnasialen Oberstufe. International stellt die Arbeit von Bryk, Sebring, Allensworth, Luppescu und Easton (2010) zur Schulreform der Jahre 1988-1996 in Chicago ein wegweisendes Beispiel dar.

Schulentwicklung im engeren Sinn kann definiert werden als die „bewusste und systematische Entwicklung von Einzelschulen“ (Rolff, 1998, S. 326). Diese Art der Schulentwicklung wird von Rolff als *intentionale* Schulentwicklung oder Schulentwicklung 1. Ordnung bezeichnet. In der Literatur finden sich verschiedene Ansätze zur Schulentwicklung, die aus unterschiedlichen Perspektiven und Bezugstheorien hervorgehen (Bohl, 2009; Rahm, 2005). Schulentwicklung findet auf verschiedenen Ebenen statt, so wird Schulentwicklung von Rolff als „Trias von personaler Entwicklung, Unterrichtsentwicklung und Organisationsentwicklung“ beschrieben (1998). Für substantielle Veränderungen müssen Veränderungen auf einer Ebene zu Veränderungen auf den anderen Ebenen führen, so dass der Ansatzpunkt der Schulentwicklung nicht festgelegt ist (Rolff, 2010). In diesem Beitrag interessieren uns v.a. die Aktivitäten auf Ebene der Organisation und die Bedeutung der Organisationsentwicklung für die Entwicklung von Schulen.

1.4 Prädiktoren der Entwicklung von Schulen

In der Schulentwicklungsforschung muss, wie bereits angesprochen, die Relevanz einzelner Prädiktoren nicht mit jener in der Schulqualitätsforschung übereinstimmen. Während im Bereich der Schulqualitätsforschung nach der „These der primären Bedeutung proximaler Faktoren“ (siehe Abschnitt 1.2) die schülernahen Faktoren den größten Effekt auf die individuelle Schülerleistung haben, könnten für die Entwicklung von Schulen gerade die sogenannten distalen Faktoren, die Faktoren auf Schulebene, eine besondere Bedeutung besitzen, denn Veränderungen der Schule als Organisation müssen sich auf die Organisation selbst beziehen und von Kollegien und Schulleitungen getragen werden. Systemtheoretisch umformuliert: Was für das Lehr-Lern-System proximal ist, ist für das System Schule als Organisation distal, und umgekehrt.

Im Folgenden stellen wir die schon angesprochenen Faktoren, die individuelle Förderung, die ganztägige Schulorganisation und Evaluationsaktivitäten, dar.

Individuelle Förderung

Eine wirksame individuelle Förderung von Schülerinnen und Schülern wurde in den letzten Jahren verstärkt eingefordert. Dies findet unter anderem in einer stärkeren Verankerung in den rechtlichen Richtlinien Ausdruck. So ist beispielsweise in Nordrhein-Westfalen „mit Verabschiedung des neuen Schulgesetzes im Jahr 2006 die individuelle Förderung jeder Schülerin und jedes Schülers als Leitidee nordrhein-westfälischer Bildungspolitik erstmals verankert worden“ (Ministerium für Schule und Weiterbildung des Landes Nordrhein-Westfalen, 2010, S. 3). Auch in den Richtlinien anderer Bundesländer wird die hohe Relevanz individueller Förderung mehr oder weniger deutlich hervorgehoben.

Allgemein bedeutet individuelle Förderung nach Fischer, Mönks und Westphal (2008) „die Anpassung des Forder-Förder-Angebots der vorschulischen und schulischen Umwelt an die kognitiven, sozial-emotionalen und psycho-motorischen Forder-Förder-Bedürfnisse des Kindes und Jugendlichen mit dem Ziel seiner optimalen Begabungsentfaltung und Persönlichkeitsentwicklung“ (S. 1). Individuelle Förderung der Schülerinnen und Schüler kann auf vielfältige Weise erreicht werden, und die Art der Umsetzung wird in den rechtlichen Richtlinien nicht festgelegt. Klieme und Warwas (2011) unterscheiden drei Varianten, die in der Schulpraxis in Deutschland verbreitet sind: (a) Trainings- und Zusatzangebote, z.B. für den Spracherwerb, (b) vielfältige, offene Lerngelegenheiten und (c) adaptiven Unterricht. Empirische Ergebnisse deuten darauf hin, dass Training und adaptiver Unterricht dann lernwirksam sind, wenn sie lernpsychologisch und didaktisch fundiert sowie gut strukturiert eingesetzt werden, während vielfältige Angebote und offene Lerngelegenheiten, wenn überhaupt, eher im affektiven Bereich wirken.

Adaptiver Unterricht wurde in PISA 2009 für Deutschland anhand einer Skala zu Differenzierungsmaßnahmen im Deutschunterricht untersucht (Hertel et al., 2010, S. 134-135), aber nicht in PISA 2000. In der vorliegenden Studie beschränken wir uns daher auf extracurriculare Angebote sowie die Wahrnehmung von Förderkursen in Deutsch durch die Schülerinnen und Schüler und prüfen, ob der Umfang extracurricularer Angebote einen Nutzen für die Entwicklung des Leseleistungsniveaus und der Beziehungsqualität hat. Angenommen wird, dass das Ausmaß solcher Angebote über den Zeitraum von neun Jahren keinen oder nur einen geringen Einfluss auf die Kriterien für die Entwicklung der Schulen hat, weil es in der Regel wenig mit den oben genannten Kernprozessen der schulischen Organisation (z.B. Entwicklung von Schulcurriculum und Leistungserwartungen; Zusammenarbeit zwischen Schulleitung, Kollegium und Eltern; Evaluationspraxis) verbunden ist.

Ganztägige Schulorganisation

Seit Anfang des neuen Jahrhunderts bildet der Ausbau von Ganztagsschulen einen Schwerpunkt der Bildungspolitik in Deutschland. Die ersten PISA-Ergebnisse lieferten dafür Argumente, aber die Ziele sind vielfältig. So sieht das Forum Bildung (Arbeitsstab Forum Bildung, 2001) in Ganztagsschulen eine Möglichkeit der individuellen Förderung, da man den Schülerinnen und Schülern hier sowohl zeitlich als auch erzieherisch besser gerecht werden kann als in Halbtagschulen (Klemm, 2011; vgl. auch Rabenstein, 2009). Verschiedene Funktionen von Ganztagschule werden auch bei Holtappels (2006, S. 9-11) deutlich; u.a. werden „ganztägige Schulformen als Reaktion auf gewandelte Bildungsanforderungen“ verstanden und die „erweiterte Schulzeit zur Entwicklung der Lernkultur und der Förderungsintensität“ genutzt, häufig aber steht der Betreuungsbedarf berufstätiger oder alleinerziehender Eltern im Vordergrund.

Die bundesweite Studie zur Entwicklung von Ganztagsschulen (Klieme, Fischer et al., 2010) zeigt, dass fach- und lernbezogene Angebote deutlich seltener vorkommen als sportliche, musische oder Freizeit-Angebote, und dass die individuelle Nutzung oft sporadisch ist. Individuelle Wirkungen sind eher schwach, aber vor allem im affektiven Bereich und im Sozialverhalten statistisch nachweisbar. Hinsichtlich ihrer Schulleistungen profitieren teilnehmende Schülerinnen und Schüler, sofern die Qualität der Angebote gut ist. Die Einführung des Ganztagsbetriebs ist zunächst eine schulorganisatorische Maßnahme: Es müssen Konzepte entwickelt und in Gremien verabschiedet, Räume und Zeitabläufe neu geplant, externe Partner gefunden, Finanzen verwaltet und Personal eingestellt werden. Wir untersuchen, inwieweit diese Veränderungen der Schulorganisation einen Beitrag zur Entwicklung von Schulen leisten.

Evaluationsaktivitäten

Evaluationsaktivitäten können als ein distaler Faktor aufgefasst werden, da sie meistens von den Lernaktivitäten der einzelnen Schülerin/des einzelnen Schülers weiter entfernt auf der Ebene der Schule angesiedelt sind. Dennoch ist die „Implementierung systematischer interner und externer Evaluationsverfahren“ nach van Ackeren und Brauckmann (2010) eine der „wahrscheinliche[n] Ursachen für den Erfolg von Bildungssystemen“ (S. 42). Klieme (2005) sieht Evaluation als eine „Abfolge von teilweise einander bedingenden Entscheidungen“. Evaluation ist „der Prozess, innerhalb dessen eine zweckgerichtete Auswahl von Bewertungskriterien erfolgt, eine Institution oder Maßnahme auf Basis dieser Kriterien systematisch untersucht und bewertet wird und eine Kommunikation über die Bewertung stattfindet mit dem Ziel, Konsequenzen abzuleiten“ (S. 41).

Evaluation lässt sich auf unterschiedliche Art klassifizieren (vgl. Holtappels, 2003). Unterteilt man Evaluation nach den durchführenden Personen, unterscheidet man *interne* Evaluation bzw. Selbstevaluation, bei der die Beteiligten/die Schulen die Evaluation selber durchführen, von der *externen* Evaluation, die durch externe Personen/Institutionen durchgeführt wird. Evaluation von Schule hat gerade in den letzten Jahren stark an Bedeutung gewonnen (Boller, 2009) und wird immer stärker gesetzlich verankert (Altrichter & Maag Merki, 2010; Berkemeyer & Müller, 2010; Maag Merki, 2009).

Da die Evaluationsmaßnahmen meist unmittelbar auf Veränderungen auf der Ebene von Schulen abzielen, lässt sich hier ein stärkerer Einfluss auf die Entwicklung von Schulen vermuten.

2. Zielsetzung und Design des Schulpanels

Ziel des Schulpanels ist die Erfassung der Entwicklung der beteiligten 59 Gymnasien über einen Zeitraum von knapp einem Jahrzehnt, einschließlich der Möglichkeit, Stabilität und Veränderung in ausgewählten schulischen Merkmalen zu untersuchen. Daneben können aufgrund des längsschnittlichen Designs Hinweise auf mögliche Wirkungszusammenhänge zwischen den untersuchten Schulmerkmalen gewonnen werden. Die Abbildung von Veränderungsprozessen auf Schulebene ist in querschnittlichen Erhebungen, wie sie in der Schuleffektivitätsforschung vorwiegend genutzt werden, nicht möglich, weswegen auf Basis dieser Studien keine kausalen Rückschlüsse gezogen werden können. Klieme und Steinert (2008, S. 229-230) beschreiben zwar erste Studien, in denen „eine Verbindung des Schulentwicklungsdiskurses und quantitativ-empirischer Schuleffektivitätsforschung erkennbar“ ist, machen jedoch den Mangel an längsschnittlichen Untersuchungen auf Schulebene deutlich, die nicht nur die Ausgangsbedingungen der Schule (Input) und Ergebnisse der Schülerinnen und Schüler (Output), sondern auch die Prozesse an der Schule berücksichtigen. Die vorliegende Studie soll einen Beitrag zu dieser Form der Schulentwicklungsforschung darstellen.

Die Datenerhebung für das Schulpanel erfolgte zu zwei Messzeitpunkten in einem Abstand von neun Jahren: Die Daten aus der Teilnahme der 59 Gymnasien an PISA 2000 wurden ergänzt durch eine wiederholte Untersuchung dieser Schulen mittels der Fragebögen und Leistungstests aus PISA 2009 (den Schülerinnen und Schülern wurde eine leicht modifizierte Version des Fragebogens vorgegeben, siehe Abschnitt 4.2). Neben quantitativen Daten, die in beiden Jahren erhoben wurden und den Schwerpunkt des Designs ausmachten, wurden im Frühjahr 2010, also zehn Jahre nach der ersten Untersuchung, mittels retrospektiver Interviews qualitative Daten zu Veränderungen an den Schulen gewonnen.

3. Fragestellungen

Mit Hilfe der gewonnenen Daten sollten folgende Fragen untersucht werden: Welche Veränderungen zeigen die Schulen hinsichtlich ihrer Entwicklungsaktivitäten? Wie hängen diese Veränderungen mit Veränderungen im Leistungsniveau der Schulen bzw. Veränderungen in der wahrgenommenen Beziehungsqualität an den Schulen von 2000 zu 2009 zusammen? Daraus ergaben sich drei Hypothesen zur Veränderung von Schulmerkmalen, welche sich aus der veränderten gesetzlichen Grundlage und veränderten bildungspolitischen Prioritäten ergeben (siehe Abschnitt 1), und drei Hypothesen zum Zusammenhang mit dem Leistungsniveau bzw. der Beziehungsqualität an den Schulen,

die auf theoretischen Überlegungen beruhen (ebenfalls Abschnitt 1). Diese werden im Folgenden noch einmal gebündelt dargestellt:

1. Schülerinnen und Schüler nehmen vermehrt an Kursen zur individuellen Förderung teil.
2. Es findet eine verstärkte Ausrichtung der Schulen zu ganztägigen Lerngelegenheiten statt.
3. An den Schulen zeigt sich ein deutlicher Zuwachs von interner und externer Evaluation.
4. Die Schulorganisation als Ganztagschule beeinflusst die Veränderung der wahrgenommenen Beziehungsqualität positiv.
5. Das Niveau der individuellen Förderung einzelner Schülerinnen und Schüler (Wahrnehmung von Kursen zur individuellen Förderung; Breite des extracurricularen Angebots) steht in keiner bzw. einer geringfügigen Beziehung mit Veränderungen im Leseleistungsniveau und der wahrgenommenen Beziehungsqualität der Schulen.
6. Veränderungen hin zu mehr Evaluationsaktivitäten haben einen praktisch relevanten positiven Effekt auf die Entwicklung der Schule, insbesondere auf das Leistungsniveau im Lesen an den Schulen.

4. Methoden²

4.1 Stichprobe

Die Zielpopulation des Schulpanels bestand aus Gymnasien aus ganz Deutschland. Aus den Gymnasien, die bereits an PISA 2000 teilgenommen hatten, wurde nach Bundesland stratifiziert eine zufällige Stichprobe von 60 Schulen gezogen; im Falle fehlender Teilnahmebereitschaft am Follow-up wurde, ebenfalls nach Zufall, eine Ersatzschule bestimmt. Zu einer Teilnahme an der weiterführenden Erhebung der Fragebogen- und Leistungsdaten erklärten sich schließlich 59 der insgesamt 90 angeschriebenen Gymnasien (inklusive Ersatzschulen) bereit. In der realisierten Stichprobe sind alle Bundesländer außer Bremen vertreten.

Die Stichproben der Schülerinnen und Schüler innerhalb der Schulen bestand in 2000 aus max. 28 ($M = 25.3$) Fünfzehnjährigen (Geburtsdatum zwischen dem 1. Februar 1984 und dem 31. Januar 1985) und in 2009 aus max. 25 ($M = 21.6$) Fünfzehnjährigen (Geburtsjahr 1993). Insgesamt ergab sich in 2000 eine Fallzahl von $N = 1493$ und in 2009 von $N = 1273$ Schülerinnen und Schülern. Die um einen Monat abweichenden Zeiträume beim Geburtsdatum der Schülerinnen und Schüler ergaben sich durch ein verändertes Er-

2 Wir danken dem IEA Data Processing and Research Center (IEA DPC) für die Datenerhebung, -erfassung und -kodierung sowie für die Herstellung der Schulkontakte. Claus Carstensen (Universität Bamberg) danken wir für die Skalierung der Leistungsdaten.

hebungszeitfenster. Im Jahr 2000 fand die Erhebung der PISA-Daten im Mai und Juni statt, in 2009 einen Monat früher im April und Mai. Durch die entsprechend angepassten Zeiträume für das Geburtsdatum der Schülerinnen und Schüler ergaben sich für beide Messzeitpunkte vergleichbare Altersverteilungen der Stichproben. Den Schulleiterfragebogen füllten in 2000 $N = 55$ Schulleiterinnen und Schulleiter und in 2009 $N = 56$ Schulleiterinnen und Schulleiter aus. Interviewdaten liegen für $N = 47$ Schulen vor.

4.2 Instrumente

Die quantitative Datenbasis des Schulpanels beruht auf den Fragebögen und Leistungstests aus PISA 2000 und PISA 2009. Eine Ausnahme stellt der Schülerfragebogen dar, der für das Schulpanel in einer vom PISA-Schülerfragebogen 2009 abweichenden Version vorgegeben wurde, um eine höhere Übereinstimmung mit den Fragen in PISA 2000 zu erreichen. Die Vergleichbarkeit der Leistungsdaten wurde durch die Verwendung von Link-Items zur Reading Literacy in beiden Jahren möglich. Während die Fragebögen und Leistungstests eine Momentaufnahme der jeweiligen Messzeitpunkte darstellten, wurden in den retrospektiven leitfadengestützten Interviews der qualitativen Erhebung Veränderungen an der Schule von 2000 bis 2009 aus der Sicht einer Vertreterin bzw. eines Vertreters der (erweiterten) Schulleitung (bzw. für vier Schulen aus der Sicht einer von der Schulleitung benannten Lehrkraft) erfragt.

Lesekompetenz

Lesekompetenz wird hier entsprechend der Definition der Hauptstudie PISA 2009 aufgefasst: „Lesekompetenz wird im Sinne einer Basiskompetenz verstanden, von der angenommen wird, dass sie in modernen Gesellschaften für eine befriedigende Lebensführung in persönlicher und wirtschaftlicher Hinsicht sowie für eine aktive Teilnahme am gesellschaftlichen Leben notwendig ist. Dies beinhaltet eine funktionale Sicht auf Lesekompetenz als basales Kulturwerkzeug. Die Beherrschung der Muttersprache in Wort und Schrift gehört in allen modernen Gesellschaften zum Kernbestand kultureller und sprachlicher Literalität“ (Naumann, Artelt, Schneider & Stanat, 2010, S. 24). Lesekompetenz umfasst „understanding, using, reflecting on and engaging with written texts, in order to achieve one's goals, to develop one's knowledge and potential, and to participate in society“ (Artelt, Naumann & Schneider, 2010, S. 74; OECD, 2009, S. 23).

Für die Ermittlung des Leistungsniveaus im Lesen wurden für das Schulpanel 36 Link-Items verwendet, die sowohl in PISA 2000 als auch 2009 vorgegeben wurden. Um Unterschiede im Testheftdesign zwischen beiden Zyklen zu kontrollieren, wurden im Schulpanel nicht nur die Testhefte aus der internationalen PISA-2009-Studie verwendet, sondern zusätzlich auch Testhefte in der Zusammenstellung aus PISA 2000. Auswirkungen der Unterschiede im Testheftdesign auf die Itemschwierigkeiten wurden im Skalierungsmodell berücksichtigt und statistisch kontrolliert. Für die Ermittlung des Lesekompetenzniveaus in den Schulen lagen im Schnitt für 2000 pro Schule

$N = 19.56$ Leistungswerte vor ($Min = 9$; $Max = 22$). Aus der Erhebung 2009 liegen im Mittel $N = 14.39$ Leistungswerte pro Schule vor ($Min = 6$; $Max = 19$).

Da die Lesekompetenz der Schülerinnen und Schüler über die Link-Items der Testhefte aus 2000 und 2009 ermittelt wurde und bedingt durch das Testheftdesign nicht jede Schülerin/jeder Schüler Link-Items zur Erfassung ihrer/seiner Lesekompetenz bearbeitet hat, ist die Gesamtzahl der Leistungswerte pro Schule niedriger als die Gesamtzahl der in der Schule getesteten und befragten Schülerinnen und Schüler.

Zur Schätzung der Lesekompetenz wurden die Daten der Schülerinnen und Schüler aus den Jahren 2000 und 2009 gemeinsam skaliert, die Itemschwierigkeiten wurden frei geschätzt. Damit sind die resultierenden Werte nicht auf der internationalen PISA-Skala verankert, aber optimal für den Vergleich des Leistungsniveaus zwischen 2000 und 2009. Für die Analysen wurden Punktschätzer (Weighted Likelihood Estimates, WLEs) der individuellen Lesekompetenz verwendet.

Wahrnehmung der Beziehungsqualität durch die Schülerschaft

Die Skala zur Beziehungsqualität gibt an, inwiefern sich Schülerinnen und Schüler durch ihre Lehrpersonen gut behandelt fühlen. Die wahrgenommene Beziehungsqualität wurde als ein Aspekt von Schülerwohlbefinden über drei Items aus dem Schülerfragebogen erfasst. Die Items erforderten eine Antwort im Antwortrange von „stimme überhaupt nicht zu“ (1) bis „stimme ganz zu“ (4) auf Aussagen wie „Die meisten meiner Lehrer/innen interessieren sich für das, was ich zu sagen habe“. Im Erhebungsjahr 2000 hat diese Skala eine Reliabilität von $\alpha = .68$ und in 2009 von $\alpha = .71$. Ein höherer Wert auf der Skala steht für eine positivere wahrgenommene Beziehungsqualität.

Individuelle Förderung

Als Indikatoren der individuellen Förderung durch die Schulen wurden die Breite des extracurricularen Angebots und die Teilnahme der Schülerschaft an Förderangeboten berücksichtigt.

Breite des extracurricularen Angebots

Sowohl im Jahr 2000 als auch im Jahr 2009 wurden im PISA-Schulleiterfragebogen außerunterrichtliche Angebote erfasst. Die verwendeten Items unterscheiden sich für beide Messzeitpunkte erheblich, so dass die Angaben nur eingeschränkt zueinander in Beziehung gesetzt werden können. Als Indikator des extracurricularen Angebots einer Schule wird zu beiden Erhebungszeitpunkten der prozentuale Anteil der von der Schule angebotenen extracurricularen Optionen an den im Schulleiterfragebogen erfragten extracurricularen Optionen verwendet. Im Jahr 2000 gaben Schulleiterinnen und Schulleiter für 35 extracurriculare Angebote an, ob diese an ihrer Schule stattfinden. Für 29 der Items lagen Ja/Nein-Antworten vor; für sechs Items wurden die Häufigkeitsangaben der Schulleiter von 1 (nie) bis 4 (mehr als viermal im Jahr) umkodiert zu Ja/Nein-Antworten. Im Jahr 2009 bejahten oder verneinten die Schulleiter das Bestehen von 22 ex-

tracurricularen Angeboten. Zu den Angeboten, die zu beiden Erhebungszeitpunkten erfasst wurden, gehören beispielsweise Leseklubs, musisch-künstlerische Angebote, mathematisch-naturwissenschaftliche Angebote und eine Schülerzeitung. Die Frage nach dem Angebot eines Debattierklubs hingegen wurde nur im Erhebungsjahr 2009 gestellt, während die Frage nach Präsentationen von Verlagen oder Buchhandlungen spezifisch für das Jahr 2000 war. Der Index zur Breite des extracurricularen Angebots (Prozentanteil der vom Schulleiter bejahten Angebote) hat in 2000 eine Reliabilität von $\alpha = .70$ und in 2009 von $\alpha = .74$.

Teilnahme der Schülerinnen und Schüler an Deutsch-Förderangeboten

Neben der Bereitstellung von extracurricularen Angeboten interessiert die Nutzung von Förderangeboten durch die Schülerinnen und Schüler. Da im vorliegenden Beitrag die PISA-Lesekompetenz als Leistungsmaß verwendet wurde, wird die folgende Frage nach der Teilnahme an Förder-/Stützkursen in Deutsch mit den Antwortkategorien „Nein, nie“ (0), „Ja, gelegentlich“ (1) und „Ja, regelmäßig“ (2) berücksichtigt: „Hast du in den letzten 3 Jahren in der Schule zusätzlichen Unterricht gehabt, um deine Leistungen zu verbessern?“.

Ganztägige Schulorganisation

Fragen zur ganztägigen Schulorganisation wurden aufgrund des Fehlens entsprechender Items im PISA-2000-Schulleiterfragebogen im Rahmen des retrospektiven Interviews gestellt. Eine Schule wurde in Anlehnung an die Definition der Kultusministerkonferenz (KMK, 2011, S. 9) als Ganztagsschule definiert, wenn an mindestens drei Tagen in der Woche ein ganztägiges Angebot besteht, dieses mindestens sieben Zeitstunden umfasst und Mittagsverpflegung für die Schülerschaft angeboten wird. Eine Schule wird als Ganztagsschule kodiert (1), wenn sie diese drei Kriterien erfüllt. Treffen weniger als die drei Kriterien zu, ist die Schule nach dieser Kodierung keine Ganztagsschule (0). Das Kriterium der konzeptuellen Verbindung von Unterricht und Ganztagsangebot, das die KMK zusätzlich bei der Kategorisierung einer Schule als Ganztagsschule voraussetzt, konnte hier nicht berücksichtigt werden; diese Forderung bestand im Jahr 2000 noch nicht.

Evaluation

Evaluationsaktivitäten an den Schulen wurden im Rahmen der retrospektiven Interviews erhoben. Hier wurden die Interviewpartner nach Veränderungen sowohl in der internen als auch in der externen Evaluation befragt. Aus diesen Angaben wurde jeweils eine Variable zum Vorhandensein von Evaluation zwischen 2000 und 2009 erzeugt. Für Schulen, für die im Interview ein konstantes oder ein vermehrtes Ausmaß an interner bzw. externer Evaluation im Zeitraum 2000 bis 2009 angegeben wurde, wurde Vorhandensein von interner bzw. externer Evaluation (1) kodiert. Bei Schulen, für die eine Abnahme der Evaluationsaktivitäten bzw. kein Stattfinden von Evaluationsaktivitäten berichtet wurde, wurde Fehlen interner bzw. externer Evaluation (0) kodiert. Da eine

Abnahme von Evaluationsaktivitäten weder für interne noch für externe Evaluation berichtet wurde, enthält die Fehlen-Kategorie ausschließlich Schulen, die keine Evaluationsaktivitäten berichteten. Für beide Prädiktoren wurde demnach – anders als bei den anderen Schulmerkmalen, für die messzeitpunktspezifische Angaben vorlagen – nur jeweils eine Variable erzeugt, die das Vorhandensein von Evaluationsaktivitäten im Zeitraum 2000 bis 2009 erfasste.

Zusammensetzung der Schülerschaft

Die Zusammensetzung der Schülerschaft wurde hinsichtlich zweier Komponenten untersucht: soziale Komposition und Migrantenanteil. Die soziale Komposition der jeweiligen Schule ergab sich durch den mittleren sozioökonomischen Status der Familien ihrer Schülerinnen und Schüler. Die Angaben der Eltern und Schülerinnen und Schüler zum Elternberuf wurden mit Hilfe der International Standard Classification of Occupation (ISCO-88) klassifiziert und in den internationalen sozioökonomischen Index (International Socio-Economic Index of Occupational Status, ISEI) umgerechnet (vgl. Ganzeboom & Treiman, 1996). Der sozioökonomische Index wurde anhand der Elternangaben bestimmt, bei fehlenden Werten wurden die Angaben aus dem Schülerfragebogen herangezogen. Auf dieser Basis wurde für die Schülerinnen und Schüler der HISEI (highest ISEI), der höchste sozioökonomische Index der Familie, gebildet, indem der jeweils höchste Wert aus Vater- und Mutter-ISEI ausgewählt wurde.

Eine Schätzung für den Anteil an Migranten wurde je Schule aus den Schülerangaben zum Geburtsort der Schülerin bzw. des Schülers, ihrer/seiner Mutter und ihres/seines Vaters ermittelt. Der Migrationsstatus wurde hier gemäß dem Vorgehen bei PISA 2009 definiert. Eine Schülerin/ein Schüler gilt dabei als Schüler/in mit Migrationshintergrund, wenn mindestens ein Elternteil im Ausland geboren wurde. Eine genaue Darstellung der Definition kann Stanat, Rauch und Segeritz (2010, S. 203-204) entnommen werden.

4.3 Analysemethoden

Angesichts der hierarchischen Struktur der Daten (Schüler geschachtelt in Schulen) erscheint für die Untersuchung der Stabilität der Schulmerkmale bzw. für die Vorhersage von Veränderungen in den Schulmerkmalen die Analyse mithilfe von Mehrebenenmodellen als die angemessene Vorgehensweise. Ein Einsatz gängiger längsschnittlicher Mehrebenenmodelle – als Mehrebenen-Regressionsmodell (z.B. Singer & Willett, 2003) oder in Form von Mehrebenen-Pfad- bzw. Strukturgleichungsmodellen (z.B. Kaplan, 2008) – ist im vorliegenden Fall allerdings nicht möglich, da solche Modelle wiederholte Erhebungen mit denselben Schülerinnen und Schülern voraussetzen. Eine Messwiederholung liegt jedoch nur auf der Ebene der Schulen vor. Eine naheliegende Lösung bestünde darin, die Daten messzeitpunktspezifisch zu aggregieren und auf Mehrebenenmodelle zugunsten einer Auswertung mit den Schulen als Analyseeinheiten zu verzichten. Insbesondere im Falle kleiner Gruppen (wenige Schüler je Schule)

und geringer Intraklassenkorrelationen (geringer Varianzanteil auf Schulebene) könnte es mit dieser Methode zu fehlerhaften Resultaten kommen, und es sollte auf alternative mehr Ebenenanalytische Auswertungsstrategien zurückgegriffen werden.

Um die Stabilität von Schulmerkmalen sowie die wechselseitige längsschnittliche Beziehung zwischen Schulmerkmalen in PISA 2000 und PISA 2009 zu untersuchen, wurden Mehrebenen-Pfadmodelle verwendet, die für eine Messwiederholung auf der Schulebene angepasst wurden. Dazu wurden Modelle mit einer Schüler- und einer Schulebene spezifiziert; die Ebene der Schulklasse konnte nicht berücksichtigt werden, da die PISA-Stichproben der Fünfzehnjährigen nur auf Basis ihrer Schul- und nicht auf Basis ihrer Klassenzugehörigkeit rekrutiert wurden. Der Typus der verwendeten Modelle wird exemplarisch in Abbildung 1 veranschaulicht. Der untere Teil der Abbildung stellt die Schülerebene (Ebene 1), der obere Teil die Schulebene dar (Ebene 2). Die Pfeile stehen für die angenommenen gerichteten oder ungerichteten Zusammenhänge zwischen den Variablen. Zur Untersuchung der wechselseitigen Beziehung zwischen Schulmerkmalen wurden auf Schulebene Cross-lagged-panel-Modelle spezifiziert (vgl. Abbildung 1). In diesen Modellen wurden jeweils ein Kriterium (Leseleistung oder Beziehungsqualität) und ein Prädiktor berücksichtigt. Zu diesen Prädiktoren zählten das Ganztagsangebot, der Index zur Breite des extracurricularen Angebots, die Teilnahme der Schülerinnen und Schüler an Deutsch-Förderkursen, der HISEI, der Migrantenanteil und das Vorhandensein von interner bzw. externer Evaluation. Die Stabilität der Schulmerkmale wurde in Modellen analysiert, in die jeweils nur das interessie-

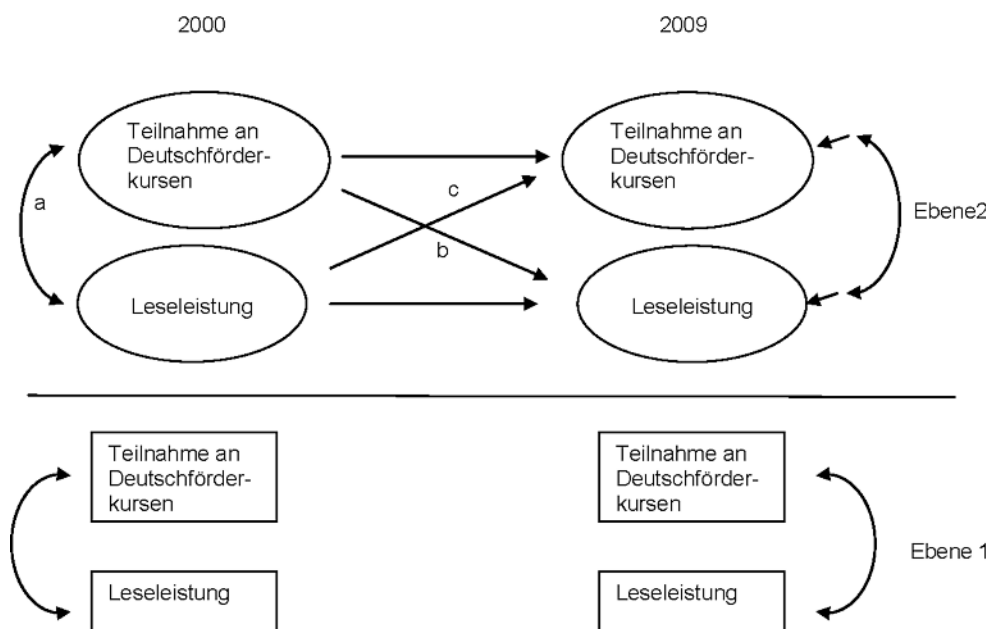


Abb. 1: Schematische Darstellung der Cross-lagged-panel-Modelle am Beispiel der Leseleistung und der Teilnahme an Deutsch-Förderkursen durch die Schülerinnen und Schüler

rende Merkmal einging. Die Stabilität wurde dann über die Korrelation des Merkmals auf Schulebene zwischen PISA 2000 und 2009 ermittelt.

Einen Sonderfall unter den auf Schulebene einbezogenen Variablen stellt die Evaluation dar. Hierfür liegen keine unabhängig in 2000 und 2009 erhobenen Daten vor (siehe Abschnitt 4.2, Evaluation). In die Analysen ging folglich nur eine Variable „Vorhandensein interner Evaluation“ bzw. „Vorhandensein externer Evaluation“ als Prädiktor für die Kriterien in 2009 ein.

Im Falle von Merkmalen, die auf Schülerebene erhoben wurden, wurden auf Schülerebene lediglich die Varianzen und – bei zwei Merkmalen dieses Typs in einem Modell (vgl. Abbildung 1) – die Kovarianzen innerhalb desselben Messzeitpunkts frei geschätzt. Alle Kovarianzen zwischen den Zeitpunkten wurden dagegen auf null gesetzt, da auf Schülerebene keine Messwiederholung stattgefunden hatte.

Die Ermittlung der Stabilität der Schulmerkmale und die Analysen zur Vorhersage von Veränderung erfolgten mithilfe von Mplus 5. Die Veränderungen im Niveau der Schulmerkmale zwischen 2000 und 2009 wurden mit SPSS 18 analysiert.

5. Ergebnisse

Im Folgenden werden die Ergebnisse zur Veränderung im Niveau der Schulqualitätsmerkmale und unserer Kriterien für die Entwicklung von Schulen zwischen 2000 und 2009 dargestellt (Abschnitt 5.1 und 5.2). In Abschnitt 5.3 wird die Stabilität der Schulmerkmale berichtet, in 5.4 wird dann auf die Vorhersage der Veränderung der Kriterien eingegangen.

5.1 Veränderungen im Niveau der Schulqualitätsmerkmale

Veränderung in der Förderpraxis der Schulen

Ein t-Test für abhängige Stichproben zeigt, dass Schülerinnen und Schüler vermehrt an Förderkursen in Deutsch teilnehmen ($t [58] = 3.142, p = .003$). Der Anteil der Schülerinnen und Schüler, die mindestens gelegentlich an Förderkursen in Deutsch teilnehmen, ist in 2009 mit 5% der Schülerinnen und Schüler, für die Angaben zur Teilnahme vorliegen, weiterhin sehr gering, jedoch signifikant höher als in 2000 (2%).

Veränderung der Schulorganisation durch Ganztagsangebote

Es zeigt sich für unsere gymnasiale Stichprobe ein signifikanter Trend hin zur Ganztagschule (Wilcoxon-Test: $Z = -4.123, p < .001$). 17 Schulen waren 2000 nach unserer Definition keine Ganztagschulen, erfüllen im Jahr 2009 jedoch die Kriterien. Für 14 Schulen ergibt sich auf Grundlage unserer Definition keine Veränderung der Schulorganisation, hiervon erfüllten 2 bereits im Erhebungsjahr 2000 die Kriterien einer Ganztagschule, während die restlichen 12 in beiden Erhebungen die Kriterien nicht erfüllen.

Veränderung in der Evaluationspraxis der Schulen

Die Befragung der Schulleiter zu Evaluationsmaßnahmen an ihren Schulen und deren Veränderung weist deutlich auf eine Zunahme der Evaluation hin. Für Maßnahmen der internen Evaluation gaben 23 der 33 Schulleiterinnen und Schulleiter, die hierzu Aussagen machten, eine Zunahme der Evaluationsaktivitäten an (70%), während 4 Personen berichteten, dass die Evaluationsmaßnahmen an ihrer Schule auf einem konstanten Niveau geblieben sind. An 6 der befragten Schulen (18%) fand bis 2009 keine interne Evaluation statt. Ein Nachlassen der Maßnahmen interner Evaluation wurde für keine Schule berichtet.

Eine Zunahme von Maßnahmen der externen Evaluation berichteten 19 der 30 Schulleiterinnen und Schulleiter (63%), die hierzu Angaben machten. Ein konstantes Niveau externer Evaluation wurde von 3 Personen angegeben. Ein weiterhin bestehendes Fehlen von externer Evaluation wurde aus 8 Schulen berichtet (27%). Ein Nachlassen im Umfang externer Evaluation wurde nicht berichtet.

5.2 Veränderungen im Niveau der Kriterien

Veränderung im Leistungsniveau

Ein t-Test für abhängige Stichproben zeigt keinen signifikanten Unterschied im mittleren Niveau der Leseleistung der Schulen zwischen 2000 und 2009 ($t [58] = .559$, $p = .578$).

Veränderung in der wahrgenommenen Beziehungsqualität

Die mittlere wahrgenommene Beziehungsqualität in den Schulen verbessert sich von 2000 zu 2009 signifikant ($t [58] = -3.121$, $p = .003$).

5.3 Stabilität der Schulmerkmale über die Zeit

Die Stabilität der Merkmale über die Zeit wurde basierend auf aggregierten Daten (Schulmittelwerte), aber auch mit Hilfe von Mehrebenenmodellen bestimmt. Tabelle 1 gibt einen Überblick über die ermittelten Stabilitäten der Merkmale über den Zeitraum von neun Jahren. Neben den Stabilitäten werden für beide Erhebungszeitpunkte die Intraklassenkorrelationen (ICC) des jeweiligen Merkmals aufgeführt. Die ICC gibt den Anteil der Varianz in der jeweiligen Variablen an, der auf Varianz zwischen den Schulen zurückzuführen ist.

Die tabellarische Darstellung der Ergebnisse zur Stabilität von Leistung, Beziehungsqualität, sozio-ökonomischem Status, Migration und Teilnahme an Förderkursen verdeutlicht teilweise große Unterschiede in der festgestellten Stabilität zwischen den Analysemethoden (Mehrebenen vs. aggregiert). In den Analysen mit den aggregierten Daten zeigen sich generell schwächere Zusammenhänge als in den Mehrebenenanalysen.

Merkmal	Korrelation 2000 – 2009 (S.E.) (basierend auf aggregierten Daten)	Korrelation 2000 – 2009 (S.E.) (basierend auf Mehrebenen- analysen)	ICC 2000	ICC 2009
Leseleistung	0.32** (0.12)	0.84** (0.32)	0.03	0.04
Wahrgenommene Beziehungsqualität	-0.03 (0.13)	-0.08 (0.25)	0.06	0.04
Sozio-ökonomischer Status (HISEI)	0.67*** (0.07)	0.97*** (0.12)	0.07	0.09
Migration	0.66*** (0.07)	0.99*** (0.08)	0.08	0.09
Teilnahme an Deutsch-Förderkursen	0.11 (0.13)	0.41 (0.51)	0.01	0.05

Anmerkungen: ICC 2000 ist die Intraklassenkorrelation im Erhebungsjahr 2000; ICC 2009 ist die Intraklassenkorrelation im Erhebungsjahr 2009; * $p < .05$, ** $p < .01$, *** $p < .001$ (zweiseitige Testung).

Tab. 1: Stabilitäten der Merkmale von Schulen, die sowohl mehrebenenanalytisch als auch aggregiert auf Schulebene betrachtet werden können

Für die auf Schulebene erfassten Merkmale „ganztägige Schulorganisation“ und „Breite des extracurricularen Angebots“ können ebenfalls Stabilitäten berichtet werden. Im Falle beider Merkmale resultiert kein signifikanter Zusammenhang zwischen den beiden Erhebungszeitpunkten (extracurriculäres Angebot: $r = -0.01$, S.E. = 0.13, n.s.; ganztägige Schulorganisation: $r = 0.20$, S.E. = 0.17, n.s.).

5.4. Vorhersage von Leseleistung und Beziehungsqualität 2009

Im Folgenden werden die Ergebnisse der Cross-lagged-panel-Modelle zur Vorhersage der Veränderung von Leseleistung und Beziehungsqualität dargestellt. Tabelle 2 gibt einen Überblick über die Prädiktion der Leseleistung. Die Spalte „Korrelation 2000 mit Leseleistung“ gibt die Korrelation zwischen der Leseleistung in 2000 und dem entsprechenden Prädiktor wieder (vgl. Abbildung 1, Pfeil a). Die Spalte „Effekte auf Leseleistung 2009“ enthält die standardisierten Regressionskoeffizienten für die Vorhersage der Leseleistung 2009 durch den jeweiligen Prädiktor 2000 (vgl. Abbildung 1, Pfeil b), die Spalte „Effekte von Leseleistung 2000“ enthält die standardisierten Regressionskoeffi-

zienten für die Vorhersage des Prädiktors 2009 durch die Leseleistung 2000 (vgl. Abbildung 1, Pfeil c). In Tabelle 3 werden die Ergebnisse zur Prädiktion der Beziehungsqualität gezeigt. Die Darstellung erfolgt nach dem gleichen Prinzip wie in Tabelle 2 für die Leseleistung.

Die Art der Signifikanztestung wurde orientiert an den Befunden bisheriger Forschung und unseren formulierten Hypothesen gewählt. So wurden Zusammenhänge, für die keine Richtungsannahmen getroffen wurden, zweiseitig getestet. Zusammenhänge, für die eine Richtungsannahme plausibel schien, wurden einseitig getestet. So zeigt die bisherige Forschung einen positiven Zusammenhang von HISEI und Leseleistung (Ehmke & Baumert, 2007; Ehmke & Jude, 2010) und einen negativen Zusammenhang von Migrantensstatus und Leseleistung (Stanat et al., 2010; Schwippert, Hornberg, Freiberg & Stubbe, 2007; Walter & Taskinen, 2007), weswegen die Zusammenhänge im Jahr 2000 einseitig geprüft wurden. Für die Breite des extracurricularen Angebots und die Teilnahme an Förderkursen nahmen wir für das Erhebungsjahr 2000 einen negativen Zusammenhang mit der Leseleistung an, da diese Fördermaßnahmen im Regelfall

Prädiktor	Korrelation 2000 mit Leseleistung (S. E.)	Effekt auf Leseleistung 2009 (S. E.)	Effekt von Leseleistung 2000 (S. E.)
Individuelle Förderung			
Breite des extracurricularen Angebots	-0.36 (0.23)	-0.08 (0.26)	0.30 (0.20)
Teilnahme an Deutsch-Förderkursen	-0.70* (0.26)	0.29 (0.85)	0.26 (0.72)
Ganztägige Schulorganisation	-0.24 (0.27)	-0.13 (0.32)	0.01 (0.24)
Evaluationsaktivitäten			
Interne Evaluation	-----	0.49* (0.27)	-----
Externe Evaluation	-----	0.27 (0.42)	-----
Zusammensetzung der Schülerschaft			
HISEI	0.53** (0.16)	0.26 (0.28)	-0.33 (0.20)
Migration	-0.08 (0.32)	-0.41 (0.23)	0.10 (0.16)

Anmerkungen: Einseitig getestete Zusammenhänge sind kursiv, zweiseitig getestete Zusammenhänge normal dargestellt; * $p < .05$, ** $p < .01$, *** $p < .001$.

Tab. 2: Ergebnisdarstellung für die Prädiktion von Leseleistung basierend auf den Cross-lagged-panel-Modellen

remedial eingesetzt werden. Der Effekt von Evaluationsaktivitäten auf die Leseleistung in 2009 wurde aufgrund der gerichteten Annahme ebenfalls einseitig getestet. Für die Beziehungsqualität wurde – in Anlehnung an Ergebnisse aus StEG – ein positiver Effekt einer Veränderung hin zur ganztägigen Schulorganisation angenommen, entsprechend erfolgte die Testung hier ebenfalls einseitig.

Aus Tabelle 2 wird unter anderem ersichtlich, dass eine signifikante positive Korrelation zwischen dem mittleren sozio-ökonomischen Status in 2000 und der mittleren Leseleistung des gleichen Erhebungszeitpunkts besteht. Ein Effekt des mittleren sozio-ökonomischen Status im Jahr 2000 auf das mittlere Niveau der Leseleistung in 2009 unter Kontrolle der mittleren Leseleistung des ersten Messzeitpunktes lässt sich dagegen nicht nachweisen. Für die Teilnahme an Förderkursen in Deutsch ergibt sich zum ersten Erhebungszeitpunkt eine negative Korrelation mit der mittleren Leseleistung.

Prädiktor	Korrelation 2000 mit Beziehungs- qualität (S. E.)	Effekt auf Beziehungs- qualität 2009 (S. E.)	Effekt von Beziehungs- qualität 2000 (S. E.)
Individuelle Förderung			
Breite des extracurricularen Angebots	-0.06 (0.16)	-0.05 (0.19)	0.22 (0.16)
Teilnahme an Deutsch-Förderkursen	-0.04 (0.31)	0.47 (0.37)	0.12 (0.22)
Ganztägige Schulorganisation	<i>0.30* (0.16)</i>	<i>0.45* (0.27)</i>	0.05 (0.19)
Evaluationsaktivitäten			
Interne Evaluation	-----	0.40* (0.19)	-----
Externe Evaluation	-----	-0.28 (0.19)	-----
Zusammensetzung der Schülerschaft			
HISEI	-0.08 (0.23)	-0.08 (0.19)	0.03 (0.13)
Migration	-0.29 (0.22)	-0.32 (0.26)	-0.16 (0.17)

Anmerkungen: Einseitig getestete Zusammenhänge sind kursiv, zweiseitig getestete Zusammenhänge normal dargestellt; * $p < .05$, ** $p < .01$, *** $p < .001$.

Tab. 3: Ergebnisdarstellung für die Prädiktion von Beziehungsqualität basierend auf den Cross-lagged-panel-Modellen

Auch in den Cross-lagged-panel-Modellen mit der Beziehungsqualität als Kriterium resultieren einige signifikante Zusammenhänge (vgl. Tabelle 3). Zum ersten Erhebungszeitpunkt im Jahr 2000 hängt eine ganztägige Schulorganisation positiv mit der mittleren Beziehungsqualität zusammen. Dieser Effekt zeigt sich auch längsschnittlich unter Kontrolle der Beziehungsqualität in 2000. Ein Effekt der mittleren Beziehungsqualität 2000 auf die ganztägige Schulorganisation 2009 (unter Kontrolle der Schulorganisation 2000) findet sich, wie zu erwarten war, dagegen nicht.

Wurden im Zeitraum von 2000 bis 2009 interne Evaluationsmaßnahmen an den Schulen durchgeführt, wirkt sich dies positiv sowohl auf die mittlere Leseleistung 2009 als auch auf die mittlere Beziehungsqualität 2009 aus. Maßnahmen der individuellen Förderung 2000 (Breite des extracurricularen Angebots und Teilnahme an Deutsch-Förderkursen) dagegen stehen erwartungsgemäß in keiner Beziehung zur Leseleistung und Beziehungsqualität im Jahr 2009.³

6. Zusammenfassung und Diskussion

Im vorliegenden Beitrag wurde die Entwicklung der Schulen im PISA-Schulpanel hinsichtlich verschiedener Aspekte untersucht. Erstes Ziel war es, Veränderungen von bzw. Niveauunterschiede in ausgewählten Schulmerkmalen zwischen 2000 und 2009 zu bestimmen.

Bei den untersuchten Gymnasien zeigt sich die erwartete Entwicklung der Organisationsform der Schulen hin zu mehr Ganztagsschulbetrieb. Eine wichtige Rolle spielten hier sicherlich die starke Förderung des Ausbaus von Ganztagsschulen durch Gelder des Investitionsprogramms „Zukunft Bildung und Betreuung (IZBB)“ (2003 bis 2009) sowie der Bedarf der Schulen nach einer Ausweitung des Unterrichts auf den Nachmittag als Folge der Einführung des achtjährigen Gymnasiums.

Die leicht gewachsene Teilnahme der Schülerinnen und Schüler an Deutsch-Förder- und Stützkursen könnte als Indikator für ein verstärktes Förderangebot und eine erhöhte Sensibilisierung an den Schulen für individuelle Förderbedarfe nach dem schlechten Abschneiden in PISA 2000 interpretiert werden. Ein deutlicher Zuwachs zwischen 2000 und 2009 zeigte sich bei den Maßnahmen der internen Evaluation. Nur knapp 18% der befragten Interviewpartner berichteten, dass im Jahr 2009 wie bereits im Jahr 2000 keine interne Evaluation stattfand. Die restlichen Schulen berichteten ein mindestens konstantes Ausmaß an interner Evaluation oder sogar vermehrte Aktivitäten der internen Evaluation. Die untersuchten Schulen wurden demnach (freiwillig oder aufgrund äußeren Drucks) vermehrt selbst aktiv im Bemühen, Genauerer über ihre aktuelle Situation zu erfahren. Auch bei den Maßnahmen der externen Evaluation konnte

3 Da im Falle der Merkmale der individuellen Förderung von keinem Effekt auf Leseleistung und Beziehungsqualität in 2009 ausgegangen wurde, wurde bei diesen Tests statt eines Signifikanzniveaus von 5% ein Niveau von 20% angesetzt. Die Effekte sind nach diesem Kriterium nicht signifikant ($p > .20$).

eine deutliche Zunahme im zurückliegenden Jahrzehnt festgestellt werden. Allerdings gaben im Vergleich zur internen Evaluation mehr – knapp ein Viertel – der Schulen an, dass Maßnahmen der externen Evaluation weiterhin fehlten. Dieser Befund überrascht, da man erwarten kann, dass an allen Schulen externe Evaluation stattgefunden hat, z.B. in Form von Vergleichsarbeiten und Schulinspektionen. Ein Grund hierfür könnte darin liegen, dass die Schulinspektion an diesen Schulen tatsächlich noch nicht stattgefunden hat; eine weitere Erklärung für diese Diskrepanz könnte die Wahrnehmung des Konzepts der Evaluation durch die Schulleiter sein: Es wurden keine objektiven Fragen nach Vorkommen und Häufigkeit einzelner Evaluationsmaßnahmen gestellt, stattdessen wurde über eine offene Frage erfasst, welche Veränderungen in der externen Evaluation die Interviewpartner in den vergangenen zehn Jahren sehen. So wurden die Interviewpartner gefragt, ob sie Veränderungen wahrnehmen („Hat sich die Evaluationspraxis an Ihrer Schule in den letzten zehn Jahren verändert?“) und wie diese aussahen („Welche Veränderungen gab es?“). Möglicherweise gehören für viele Interviewpartner z.B. Vergleichsarbeiten nicht zum Konzept „externe Evaluation“. Dies könnte zum einen daran liegen, dass Schulen externe Evaluation oftmals mit Schulinspektion gleichsetzen, und zum anderen daran, dass von administrativer Seite gegenüber den Schulen der Unterrichtsentwicklungsaspekt von Vergleichsarbeiten stärker betont wird als der Evaluationsaspekt (Isaac, 2010; Orth, 2002; Peek, 2009; Petilliot-Becker & Müller-Rosigkeit, 2010; Ramm, 2007), so dass die Wahrnehmung der Schulleiterinnen und Schulleiter von Vergleichsarbeiten als Instrument der externen Evaluation ggfs. nicht gegeben ist.

Das mittlere Leistungsniveau der im Schulpanel untersuchten Gymnasien ist von 2000 zu 2009 konstant geblieben. Dieser Befund ist konform mit den Ergebnissen von PISA 2009 (Naumann et al., 2010), wonach sich die Leseleistung der Schülerinnen und Schüler im oberen Kompetenzbereich zwischen PISA 2000 und PISA 2009 nicht signifikant verändert hat. Die von den Schülerinnen und Schülern wahrgenommene Qualität der Beziehung zu ihren Lehrerinnen und Lehrern verbesserte sich hingegen über den Zeitraum von neun Jahren. Hiermit übereinstimmend berichteten die Schülerinnen und Schüler in PISA 2009 eine insgesamt positivere Lehrer-Schüler-Beziehung als in PISA 2000 (Hertel et al., 2010).

Zweites Ziel der Untersuchung war die Bestimmung der Stabilität der Schulmerkmale. Die Analysen auf der Grundlage von aggregierten Daten und die Mehrebenenanalysen führten zu teilweise deutlich unterschiedlichen Ergebnissen. Diese Disparitäten sind primär auf die geringen Varianzanteile auf Schulebene (ICC) und die daraus resultierende geringe Reliabilität der auf Schulebene aggregierten Schülerangaben zurückzuführen (vgl. z.B. Snijders & Bosker, 1999). Infolge dieser geringeren Reliabilität kommt es in den Analysen mit aggregierten Daten zu einer Unterschätzung der Zusammenhänge auf der Schulebene. Die Mehrebenenanalysen dagegen ermöglichen es, die Zusammenhänge basierend auf den tatsächlichen Unterschieden zwischen den Schulen zu schätzen. Die Gegenüberstellung verdeutlicht den Vorteil der mehrebenenanalytischen Vorgehensweise, wenn schulische Merkmale längsschnittlich untersucht werden sollen.

Die Befunde aus den Mehrebenenanalysen weisen bei signifikantem Ergebnis auf eine hohe Stabilität der betrachteten Merkmale hin. Dies gilt insbesondere für die sozi-

ale Zusammensetzung und den Migrantenanteil. Immerhin knapp 10% der im Jahr 2009 beobachteten Varianz des sozialen Status und des Migrationsstatus liegen zwischen den Schulen, und dieser Anteil wird durch die bereits im Jahr 2000 erhobenen Werte des Merkmals nahezu perfekt erklärt. Gymnasien unterscheiden sich also durchaus hinsichtlich der sozialen Selektion und Selbst-Selektion (durch die Eltern), und diese Unterschiede sind außerordentlich stabil. Die Varianz des „Output“, und zwar sowohl der Leseleistung als auch der wahrgenommenen Beziehungsqualität, lag in 2009 hingegen nur zu 4% auf Schulebene. Hinsichtlich der kognitiven wie der affektiven Kriterien sind sich Gymnasien also recht ähnlich. Die verbleibenden Unterschiede auf Schulebene sind im kognitiven Bereich recht stabil, im affektiven Bereich jedoch in hohem Maße instabil, was auf eine erhebliche Abhängigkeit des Schulklimas von den beteiligten Personen, z.B. von der zeitpunktspezifischen Zusammensetzung der Schülerschaft und des Kollegiums, hindeuten könnte. Die geringe Stabilität im Niveau des extracurricularen Angebots der Schulen dagegen könnte zum Teil auch auf die Unterschiedlichkeit der in 2000 und 2009 verwendeten Indizes zurückzuführen sein (siehe Abschnitt 4.2, Individuelle Förderung). Die Schulorganisation (ohne/mit Ganztagsangebot) und die Teilnahme an Förderkursen in Deutsch sind ebenfalls eher instabil. Die mangelnde Stabilität hängt sicherlich mit dem starken Ausbau von Schulen zu Ganztagschulen innerhalb des letzten Jahrzehnts zusammen. Zudem hängt die Teilnahme der Schülerinnen und Schüler an individuellen Förderkursen neben dem Bedarf der Schülerinnen und Schüler an solchen Kursen mit dem von der Schule gestellten Angebot zusammen, welches hier nicht berücksichtigt werden konnte.

Drittes Ziel unseres Beitrags war die Prädiktion der Veränderung in Leseleistung und Beziehungsqualität zwischen 2000 und 2009. Hier ließen sich einzelne Zusammenhänge mit den herangezogenen Schulqualitätsmerkmalen nachweisen. Zum einen zeigt sich der theoretisch plausible negative Zusammenhang zwischen der Teilnahme an Deutsch-Förderkursen und der Leseleistung innerhalb des Erhebungsjahres 2000. Dieser Befund kann wohl nur im Sinne eines rekursiven Effekts interpretiert werden: Schulen mit vergleichsweise niedrigem Leistungsstand im Lesen reagieren auf diese Problematik mit einem Angebot an Förderkursen. Im Längsschnitt dagegen besteht, wie von uns erwartet, kein Zusammenhang, vermutlich, da es sich bei der Teilnahme an Förderkursen primär um eine Maßnahme der individuellen Förderung der Schülerin/des Schülers und nicht um eine Maßnahme zur Entwicklung der „Organisation Schule“ handelt.

Weitere querschnittliche Zusammenhänge im Erhebungsjahr 2000 finden sich erwartungsgemäß zwischen Leistung und sozio-ökonomischem Status (vgl. Ehmke & Jude, 2010) sowie zwischen Beziehungsqualität und Ganztagsbetrieb (vgl. die StEG-Resultate; Klieme, Fischer et al., 2010). Schülerinnen und Schüler an ganztätig arbeitenden Schulen beschreiben die Beziehung zu ihren Lehrkräften im Mittel positiver als Schülerinnen und Schüler an Schulen ohne Ganztagsangebot. Dieser positive Effekt eines Ganztagsangebots auf die wahrgenommene Beziehungsqualität zeigt sich auch im Längsschnitt: Auch unter Kontrolle der wahrgenommenen Beziehungsqualität in 2000 findet sich ein positiver Effekt des Ganztagsangebots 2000 auf die wahrgenommene Beziehungsqualität in 2009. Bedeutsam für die Entwicklung sowohl der Lese-

leistung als auch der Beziehungsqualität ist nach unseren Befunden schließlich nicht die externe Evaluationspraxis, wohl aber die Nutzung interner Evaluation. Schulen, die bereits 2000 interne Evaluation betrieben und dieses Ausmaß an Evaluation beibehalten oder bis 2009 verstärkt haben, zeigen eine höhere Leseleistung und eine günstigere Beziehungsqualität. Eine mögliche Erklärung für diese Konstellation ist die stärkere Ausrichtung der Maßnahmen interner Evaluation auf die konkrete Situation an der Schule als bei externer Evaluation. Während Bildungsökonominnen, ausgerichtet an Wettbewerbsmodellen, externe Evaluation für wirksam erachten (Hanushek & Wößmann, 2010), sehen wir die Auswirkungen primär bei der internen Evaluation, weil diese infolge der Durchführung durch die Schulen selbst eine Beteiligung in der Schule voraussetzt und möglicherweise Prozesse und Aktivitäten in Gang setzt, die bei externer Evaluation nicht zwingend stattfinden.

Die untersuchten Prädiktoren wurden in Abschnitt 1.2 bzw. 1.4 mit der in der Schulqualitätsforschung üblichen Unterscheidung zwischen proximalen Merkmalen und distalen Merkmalen in Beziehung gebracht. In der vorliegenden Studie zeigte sich, dass die individuelle Förderung als proximales Merkmal des Lehr-Lerngeschehens von nur geringer, die distalen Merkmale (ganztägige Schulorganisation und interne Evaluationspraxis) dagegen von stärkerer Bedeutung für die Entwicklung der Schulen sind, insbesondere für das herangezogene affektive Kriterium. Von diesen Befunden sollte jedoch nicht auf die allgemeine Bedeutung proximaler vs. distaler Merkmale für die Schulentwicklung geschlossen werden, da der Ausschnitt der untersuchten Merkmale sehr klein ist. Dennoch legen die Ergebnisse nahe, dass eine weitere Untersuchung der relativen Bedeutung verschiedener Faktoren auf unterschiedlichen Ebenen für die Schulentwicklung unerlässlich ist.

Einschränkend ist anzumerken, dass in den bisherigen Analysen nur bivariate Zusammenhänge untersucht wurden, während multivariate Modelle in diesem Beitrag keinen Platz gefunden haben. Vertiefende Analysen unter Berücksichtigung z.B. der Schülerkomposition sind für weitere Publikationen vorgesehen.

Abschließend kann man festhalten, dass ein Zeitraum von neun Jahren, der für ein ganzes Schulsystem gerade ausreichend sein mag, um sich merklich zu verbessern (vgl. Klieme, Jude et al., 2010, S. 280), für Einzelschulen einen mehr als ausreichenden Erwartungshorizont bietet. Bryk et al. (2010) beispielsweise fanden messbare Entwicklungsunterschiede zwischen Schulen nach etwa 6 Jahren der Schulreform. Andererseits können in einem Zeitraum von fast einem Jahrzehnt viele Ereignisse die Entwicklung der Schulen beeinflusst haben, und die Annahme einer linearen Beziehung zwischen schulischen Bedingungsfaktoren und Ergebnisvariablen wird der Dynamik der Veränderung von Schulen möglicherweise nicht gerecht. Eine Möglichkeit, das Auftreten von für die Entwicklung von Schulen relevanten Ereignissen zu prüfen, liegt in einer weiteren Analyse der Interviewdaten, für die bewusst auf den Zeitraum zwischen der Erhebung 2000 und 2009 fokussiert wurde. Hier können Analysen von Einzelfällen mit besonders auffälligen Veränderungen weitere Erkenntnisse bringen. Darüber hinaus sollen in weiteren Publikationen neben den wenigen hier einbezogenen Schulmerkmalen auch

deren Qualität sowie weitere, durch Schulleiter-, Eltern- und Schülerbefragungen erhobene Schulmerkmale hinzugenommen werden.

Das Konzept eines an große Schulleistungsstudien wie PISA angehängten Schulpanels hat sich mit der vorliegenden Auswertung als fruchtbar erwiesen. Eine wichtige Chance für die Schulforschung wäre es, ein solches Design auch international vergleichend durchführen zu können, wie es bereits für PISA 2012 vorgeschlagen wurde (vgl. Klieme, Backhoff et al., 2010, S. 40).

Literatur

- Altrichter, H., & Maag Merki, K. (2010). Steuerung der Entwicklung des Schulwesens. In Dies. (Hrsg.), *Handbuch Neue Steuerung im Schulsystem* (S. 15-39). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Arbeitsstab Forum Bildung (Hrsg.) (2001). *Empfehlungen und Einzelergebnisse des Forum Bildung*. Bonn: Forum Bildung (Eigendruck).
- Artelt, C., Naumann, J., & Schneider, W. (2010). Lesemotivation und Lernstrategien. In E. Klieme, C. Artelt, J. Hartig, N. Jude, O. Köller, M. Prenzel, W. Schneider & P. Stanat (Hrsg.), *PISA 2009. Bilanz nach einem Jahrzehnt*. Münster: Waxmann.
- Baumert, J., Klieme, E., Neubrand, M., Prenzel, M., Schiefele, U., Schneider, W., Stanat, P., Tillmann, K.J., & Weiß, M. (Hrsg.) (2001). *PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich*. Opladen: Leske + Budrich.
- Baumert, J., Artelt, C., Klieme, E., Neubrand, M., Prenzel, M., Schiefele, U., Schneider, W., Tillmann, K.-J., & Weiß, M. (Hrsg.) (2003). *PISA 2000. Ein differenzierter Blick auf die Länder der Bundesrepublik Deutschland*. Opladen: Leske + Budrich.
- Berkemeyer, N., & Müller, S. (2010). Schulinterne Evaluation – nur ein Instrument zur Selbststeuerung von Schulen? In H. Altrichter & K. Maag Merki (Hrsg.), *Handbuch Neue Steuerung im Schulsystem* (S. 195-218). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Bohl, T. (2009). Theorien und Konzepte der Schulentwicklung. In S. Blömeke, T. Bohl, L. Haag, G. Lang-Wojtasik & W. Sacher (Hrsg.), *Handbuch Schule* (S. 553-559). Bad Heilbrunn: Klinkhardt/UTB.
- Boller, S. (2009). *Kooperation in der Schulentwicklung. Interdisziplinäre Zusammenarbeit in Evaluationsprojekten*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Brophy, J. (2000). *Teaching*. Brüssel/Genf: International Academy of Education/International Bureau of Education.
- Bryk, T., Sebring, P. B., Allensworth, E., Luppescu, S., & Easton, J. C. (2010). *Organizing schools for improvement. Lessons from Chicago*. Chicago: The University of Chicago Press.
- Creemers, B. P. M., & Reezigt, G. J. (1996). School level conditions affecting the effectiveness of instruction. *School Effectiveness and School Improvement*, 7(3), 197-228.
- Creemers, B. P. M., & Kyriakides, L. (2008). *The dynamics of educational effectiveness: a contribution to policy, practice and theory in contemporary schools*. London: Routledge.
- de Maeyer, S., van den Bergh, H., Rymenans, R., van Petegem, P., & Rijlaarsdam, G. (2010). Effectiveness criteria in school effectiveness studies: Further research on the choice for a multivariate model. *Educational Research Review*, 5, 81-96.
- Diedrich, M. (2008). *Demokratische Schulkultur*. Münster: Waxmann.
- Ditton, H. (2000a). Qualitätskontrolle und -sicherung in Schule und Unterricht – ein Überblick zum Stand der empirischen Forschung. *Zeitschrift für Pädagogik*, 41. Beiheft, 73-92.
- Ditton, H. (2000b). Elemente eines Systems der Qualitätssicherung im schulischen Bereich. In H. Weishaupt (Hrsg.), *Qualitätssicherung im Bildungswesen: Problemlage und aktuelle For-*

- schungsbefunde. Erfurter Studien zur Entwicklung des Bildungswesens* (S. 13-35). Erfurt: Pädagogische Hochschule.
- Ditton, H. (2007). Schulqualität. Modelle zwischen Konstruktion, empirischen Befunden und Implementierung. In J. van Buer & C. Wagner (Hrsg.), *Qualität von Schule. Ein kritisches Handbuch* (S. 83-92). Frankfurt a.M.: Peter Lang.
- Eder, F. (2002). Unterrichtsklima und Unterrichtsqualität. *Unterrichtswissenschaft*, 30, 213-228.
- Eder, F. (2006). Schul- und Klassenklima. In D. H. Rost (Hrsg.), *Handwörterbuch Pädagogische Psychologie* (S. 622-631). Weinheim/Basel: Beltz.
- Edmonds, R. (1979). Effective schools for the urban poor. *Educational Leadership*, 37(1), 15-27.
- Ehmke, T., & Baumert, J. (2007). Soziale Herkunft und Kompetenzerwerb: Vergleiche zwischen PISA 2000, 2003 und 2006. In M. Prenzel, C. Artelt, J. Baumert, W. Blum, M. Hammann, E. Klieme & R. Pekrun (Hrsg.), *PISA 2006 in Deutschland. Die Kompetenzen der Jugendlichen im dritten Ländervergleich* (S. 309-335). Münster: Waxmann.
- Ehmke, T., & Jude, N. (2010). Soziale Herkunft und Kompetenzerwerb. In E. Klieme, C. Artelt, J. Hartig, N. Jude, O. Köller, M. Prenzel, W. Schneider & P. Stanat (Hrsg.), *PISA 2009. Bilanz nach einem Jahrzehnt* (S. 277-300). Münster: Waxmann.
- Fend, H. (1998). *Qualität im Bildungswesen. Schulforschung zu Systembedingungen, Schulprofilen und Lehrerleistung*. Weinheim: Juventa.
- Fend, H. (2008). *Schule gestalten. Systemsteuerung, Schulentwicklung und Unterrichtsqualität*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Fischer, C., Mönks, F. J., & Westphal, U. (Hrsg.) (2008). *Individuelle Förderung: Begabungen entfalten – Persönlichkeit entwickeln*. Berlin: LIT.
- Fraser, B. J. (1994). Research on classrooms and school climate. In D. Gabel (Hrsg.), *Handbook of research on science teaching and learning* (S. 493-541). New York: MacMillan.
- Ganzeboom, H. B. G., & Treiman, D. J. (1996). Internationally comparable measures of occupational status for the 1988 International Standard Classification of Occupations. *Social Science Research*, 25, 201-239.
- Gerecht, M. (2010). *Schul- und Unterrichtsqualität und ihre erzieherischen Wirkungen*. Münster: Waxmann.
- Götz, T., Frenzel, A. C., & Pekrun, R. (2008). Sozialklima in der Schule. In W. Schneider & M. Hasselhorn (Hrsg.), *Handbuch der Psychologie. Band Pädagogische Psychologie* (S. 503-514). Göttingen: Hogrefe.
- Hanushek, E. A., & Wößmann, L. (2010). *The economics of international differences in educational achievement* (NBER Working Paper No. 15949). <http://www.nber.org/papers/w15949.pdf> [16.11.2012].
- Hertel, S., Hochweber, J., Steinert, B., & Klieme, E. (2010). Schulische Rahmenbedingungen und Lerngelegenheiten im Deutschunterricht. In E. Klieme, C. Artelt, J. Hartig, N. Jude, O. Köller, M. Prenzel, W. Schneider & P. Stanat (Hrsg.), *PISA 2009. Bilanz nach einem Jahrzehnt* (S. 113-151). Münster: Waxmann.
- Holtappels, H. G. (2003). *Schulqualität durch Schulentwicklung und Evaluation. Konzepte, Forschungsbefunde, Instrumente*. München: Luchterhand.
- Holtappels, H. G. (2006). Stichwort: Ganztagschule. *Zeitschrift für Erziehungswissenschaft*, 9(1), 5-29.
- Holtappels, H. G., Klemm, K., & Rolff, H.-G. (Hrsg.) (2008). *Schulentwicklung durch Gestaltungsautonomie. Ergebnisse der Begleitforschung zum Modellvorhaben 'Selbstständige Schule' in Nordrhein-Westfalen*. Münster: Waxmann.
- Isaac, K. (2010). Chancen und Grenzen von Vergleichsarbeiten – Zur schulinternen Nutzung von Daten. *SchulVerwaltung NRW*, 21(7/8), 197-199.
- Kaplan, D. (2008). *Structural equation modeling*. Thousand Oaks: Sage.

- Klemm, K. (2011). Das Bildungssystem Deutschlands. Strukturen und Strukturreformen. In H. Reinders, H. Ditton, C. Gräsel & B. Gniewosz (Hrsg.), *Empirische Bildungsforschung. Strukturen und Methoden* (S. 153-164). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Klieme, E. (2005). Zur Bedeutung von Evaluation für die Schulentwicklung. In K. Maag Merki, A. Sandmeier, P. Schuler & H. Fend (Hrsg.), *Schule wohin? Schulentwicklung und Qualitätsmanagement im 21. Jahrhundert* (S. 40-61). Zürich: Pädagogisches Institut der Universität Zürich.
- Klieme, E., Backhoff, E., Blum, W., Buckley, J., Hong, Y., Kaplan, D., Levin, H., Scheerens, J., Schmidt, W., van de Vijver, F., & Vieluf, S. (2010). *Designing PISA as a sustainable database for educational policy and research: The PISA 2012 Context Questionnaire Framework*. Paris: OECD.
- Klieme, E., Fischer, N., Holtappels, H. G., Rauschenbach, T., & Stecher, L. (2010). *Ganztagsschule: Entwicklung und Wirkungen. Ergebnisse der Studie zur Entwicklung von Ganztagsschulen 2005-2010*. http://projekt-steg.de/files/pk101111/Ergebnisbroschuere_StEG_2010-11-11.pdf [16.11.2012].
- Klieme, E., Jude, N., Baumert, J., & Prenzel, M. (2010). PISA 2000-2009: Bilanz der Veränderungen im Schulsystem. In E. Klieme, C. Artelt, J. Hartig, N. Jude, O. Köller, M. Prenzel, W. Schneider & P. Stanat (Hrsg.), *PISA 2009. Bilanz nach einem Jahrzehnt* (S. 277-300). Münster: Waxmann.
- Klieme, E., & Steinert, B. (2008). Schulentwicklung im Längsschnitt. In M. Prenzel & J. Baumert (Hrsg.), *Vertiefende Analysen zu PISA 2006* (S. 221-238). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Klieme, E., Steinert, B., & Hochweber, J. (2010). Zur Bedeutung der Schulqualität für Unterricht und Lernergebnisse. In W. Bos, E. Klieme & O. Köller (Hrsg.), *Schulische Lerngelegenheiten und Kompetenzentwicklung. Festschrift für Jürgen Baumert* (S. 231-255). Münster: Waxmann.
- Klieme, E., & Warwas, J. (2011). Konzepte der individuellen Förderung. *Zeitschrift für Pädagogik*, 57(6), 805-818.
- KMK = Sekretariat der ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland (2011). *Definitionenkatalog zur Schulstatistik 2011* (2). <http://www.kmk.org/fileadmin/pdf/Statistik/Defkat2011.pdf> [04.12.2012].
- Maag Merki, K. (2009). Evaluation im Bildungsbereich Schule in Deutschland. In T. Widmer, W. Beywl & C. Fabian (Hrsg.), *Evaluation. Ein systematisches Handbuch* (S. 157-162). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Ministerium für Schule und Weiterbildung des Landes Nordrhein-Westfalen (2010). *Die Initiative „Gütesiegel Individuelle Förderung“. Ein Leitfaden für Schulen*. Neuss: Meinke Print Media Partner.
- Mortimore, P., Sammons, P., & Thomas, S. (1994). School effectiveness and value added measures. *Assessment in Education: Principles, Policy & Practice*, 1(3), 315-332.
- Naumann, J., Artelt, C., Schneider, W., & Stanat, P. (2010). Lesekompetenz von PISA 2000 bis PISA 2009. In E. Klieme, C. Artelt, J. Hartig, N. Jude, O. Köller, M. Prenzel, W. Schneider & P. Stanat (Hrsg.), *PISA 2009. Bilanz nach einem Jahrzehnt* (S. 23-71). Münster: Waxmann.
- OECD (2009). *PISA 2009 assessment framework. Key competencies in reading, mathematics and science*. Paris: OECD.
- Opdenakker, M.-C., & van Damme, J. (2000). Effects of schools, teaching and classes on achievement and well-being in secondary education: Similarities and differences between school outcomes. *School Effectiveness and School Improvement*, 11(2), 165-196.
- Orth, G. (2002). Vergleichsarbeiten. In H.-G. Rolff & H.-J. Schmidt (Hrsg.), *Brennpunkt Schulleitung und Schulaufsicht* (S. 203-223). Neuwied: Luchterhand.
- Peek, R. (2009). Dateninduzierte Schulentwicklung. In H. Buchen & H.-G. Rolff (Hrsg.), *Professionswissen Schulleitung* (S. 1343-1367). Weinheim/Basel: Beltz.

- Petilliot-Becker, I., & Müller-Rosigkeit, E. (2010). Diagnose- und Vergleichsarbeiten – Zielsetzungen und Konzeption. *Schulverwaltung Baden-Württemberg*, 19(6), 122-124.
- Rabenstein, K. (2009). Individuelle Förderung in unterrichtsergänzenden Angeboten an Ganztagschulen – ein Fallvergleich. In S. Appel, H. Ludwig & U. Rother (Hrsg.), *Jahrbuch Ganztagschule 2010. Vielseitig fördern* (S. 23-33). Schwalbach am Taunus: Wochenschau-Verlag.
- Rahm, S. (2005). *Einführung in die Theorie der Schulentwicklung*. Weinheim/Basel: Beltz.
- Ramm, G. (2007). VERA kommt in die Sekundarstufe. *Schul-Management*, 4, 20-23.
- Rolff, H.-G. (1998). Entwicklung von Einzelschulen: Viel Praxis, wenig Theorie und kaum Forschung. Ein Versuch, Schulentwicklung zu systematisieren. In H.-G. Rolff, K.-O. Bauer, K. Klemm, H. Pfeiffer (Hrsg.), *Jahrbuch der Schulentwicklung*, Bd. 10 (S. 295-325). Weinheim: Juventa.
- Rolff, H.-G. (2010). Schulentwicklung als Trias von Organisations-, Unterrichts- und Personalentwicklung. In T. Bohl, W. Helsper, H. G. Holtappels & C. Schelle (Hrsg.), *Handbuch Schulentwicklung* (S. 29-36). Bad Heilbrunn: Klinkhardt/UTB.
- Sammons, P., Hillman, J., & Mortimore, P. (1995). *Key characteristics of effective schools: a review of school effectiveness research. A report by the Institute of Education for the Office for Standards in Education*. London: Institute of Education.
- Scheerens, J. (2008). *Review and meta-analyses of school and teaching effectiveness*. Twente: University of Twente, Department of Educational Organization and Management.
- Scheerens, J., & Bosker, R. J. (1997). *The foundations of educational effectiveness*. Oxford: Pergamon Press.
- Schwippert, K., Hornberg, S., Freiberg, M., & Stubbe, T. C. (2007). Lesekompetenzen von Kindern mit Migrationshintergrund im internationalen Vergleich. In W. Bos, S. Hornberg, K.-H. Arnold, G. Faust, L. Fried, E.-M. Lankes, K. Schwippert & R. Valtin (Hrsg.), *IGLU 2006. Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 249-270). Münster: Waxmann.
- Seidel, T. (2008). Stichwort: Schuleffektivitätskriterien in der internationalen empirischen Forschung. *Zeitschrift für Erziehungswissenschaft*, 11(3), 348-367.
- Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis*. New York: Oxford University Press.
- Snijders, T. A. B., & Bosker, R. J. (1999). *Multilevel analysis*. London: Sage.
- Stanat, P., Rauch, D., & Segeritz, M. (2010). Schülerinnen und Schüler mit Migrationshintergrund. In E. Klieme, C. Artelt, J. Hartig, N. Jude, O. Köller, M. Prenzel, W. Schneider & P. Stanat (Hrsg.), *PISA 2009. Bilanz nach einem Jahrzehnt* (S. 200-230). Münster: Waxmann.
- Steinhage, N., & Blossfeld, H.-P. (1999). *Zur Problematik von Querschnittsdaten*. <http://nbn-resolving.de/urn:nbn:de:0168-ss0ar-57590> [04.12.2012].
- Teddlie, C., & Reynolds, D. (Hrsg.) (2000). *The international handbook of school effectiveness research*. London: Falmer Press.
- Trautwein, U., Neumann, M., Nagy, G., Lüdtke, O., & Maaz, K. (Hrsg.) (2010). *Schulleistungen von Abiturienten: Die neu geordnete gymnasiale Oberstufe auf dem Prüfstand*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- van Ackeren, I., & Brauckmann, S. (2010). Internationale Diskussions-, Forschungs- und Theorienansätze zur Governance im Schulwesen. In H. Altrichter & K. Maag Merki (Hrsg.), *Handbuch Neue Steuerung im Schulsystem* (S. 41-61). Wiesbaden: VS Verlag für Sozialwissenschaften.
- van Damme, J., de Fraine, B., van Landeghem, G., Opdenakker, M.-C., & Onghena, P. (2002). A new study on educational effectiveness in secondary schools in Flanders: An introduction. *School Effectiveness and School Improvement*, 13(4), 383-397.
- van Damme, J., Opdenakker, M.-C., van Landeghem, G., de Fraine, B., Pustjens, H., & van de Gaer, E. (2006). *Educational effectiveness: an introduction to international and Flemish research on schools, teachers and classes*. Leuven: Acco.

Walter, O., & Taskinen, P. (2007). Kompetenzen und bildungsrelevante Einstellungen von Jugendlichen mit Migrationshintergrund in Deutschland: Ein Vergleich mit ausgewählten OECD-Staaten. In M. Prenzel, C. Artelt, J. Baumert, W. Blum, M. Hammann, E. Klieme & R. Pekrun (Hrsg.), *PISA 2006 in Deutschland. Die Kompetenzen der Jugendlichen im dritten Ländervergleich* (S. 337-366). Münster: Waxmann.

Anschrift der Autorin/der Autoren

Dipl. Psych. Linda Marie Bischof, Deutsches Institut für Internationale Pädagogische Forschung (DIPF), Schloßstraße 29, 60486 Frankfurt am Main, Deutschland
E-Mail: bischof@dipf.de

Dr. Jan Hochweber, Deutsches Institut für Internationale Pädagogische Forschung (DIPF), Schloßstraße 29, 60486 Frankfurt am Main, Deutschland
E-Mail: hochweber@dipf.de

Prof. Dr. Johannes Hartig, Deutsches Institut für Internationale Pädagogische Forschung (DIPF), Schloßstraße 29, 60486 Frankfurt am Main, Deutschland
E-Mail: hartig@dipf.de

Prof. Dr. Eckhard Klieme, Deutsches Institut für Internationale Pädagogische Forschung (DIPF), Schloßstraße 29, 60486 Frankfurt am Main, Deutschland
E-Mail: klieme@dipf.de

Nina Jude/Johannes Hartig/Stefan Schipolowski/Katrin Böhme/Petra Stanat

Definition und Messung von Lesekompetenz

PISA und die Bildungsstandards

1. Einleitung

Lesekompetenz ist spätestens seit der ersten Durchführung des *Programme for International Student Assessment* (PISA) im Jahr 2000 ein regelmäßig erhobener Indikator im Rahmen des internationalen und nationalen Bildungsmonitoring. Neben PISA haben im vergangenen Jahrzehnt verschiedene *Large Scale Assessments* Aussagen über das Niveau und die Verteilung der Lesekompetenz von Schülerinnen und Schülern in Deutschland getroffen. Dazu gehören beispielsweise Studien wie Deutsch Englisch Schülerleistungen International (DESI), die Internationale Grundschul-Lese-Untersuchung (IGLU bzw. PIRLS) sowie Untersuchungen, die das Institut zur Qualitätsentwicklung im Bildungswesen (IQB) auf Basis der länderübergreifenden Bildungsstandards der Kultusministerkonferenz durchführt (Bos et al., 2007; KMK, 2004a, 2004b; Klieme et al., 2008; Köller, Knigge & Tesch, 2010; Stanat, Pant, Böhme & Richter, 2012).

Auch wenn diese Studien jeweils Aussagen über den gleichen Kompetenzbereich, die Lesekompetenz, anstreben, so beruhen sie doch auf theoretisch und fachdidaktisch teils unterschiedlich fundierten Definitionen dieses Konstrukts. Unterschiede finden sich sowohl im verwendeten Testmaterial als auch in der Definition von Kompetenzstufen zur kriteriumsorientierten Beschreibung der von Schülerinnen und Schülern erreichten Kompetenzstände. Es stellt sich daher die Frage, inwieweit die in den genannten Studien erfassten Konstrukte vergleichbar sind und wie ihre Ergebnisse aufeinander bezogen werden können.

Für einen solchen Vergleich bietet es sich aktuell an, zwei Studien näher zu betrachten, die bezüglich ihrer Genese und Verwendung von Kompetenzstufenmodellen aufeinander aufbauen: PISA und der zeitgleich in der Sekundarstufe I durchgeführte Ländervergleich des IQB, der auf den Bildungsstandards basiert (Klieme et al, 2010; Köller et al., 2010).

Im Rahmen von PISA wird die Lesekompetenz von Fünfzehnjährigen in Deutschland seit dem Jahr 2000 im dreijährigen Rhythmus getestet und im internationalen Vergleich berichtet. Die erweiterten, „PISA-E“ genannten Erhebungen, die in den Jahren

2000, 2003 und 2006 durchgeführt wurden, ermöglichten es darüber hinaus, diese Kompetenzen auf der Ebene der Länder in der Bundesrepublik Deutschland darzustellen (Baumert et al., 2002; Prenzel et al., 2005, 2008).

Mit der Einführung der Ländervergleiche auf Basis der Bildungsstandards der KMK, die im Fach Deutsch für den Mittleren Schulabschluss im Jahr 2004 erfolgte (KMK, 2004a), wurde PISA-E im Jahr 2009 abgelöst. Während die internationale Erhebung von PISA 2009 weiterhin mit einer für Deutschland repräsentativen Stichprobe die Lesekompetenz von Fünfzehnjährigen erfasste, bezog sich die Kompetenzmessung auf Basis der Bildungsstandards im Jahr 2009 auf Schülerinnen und Schüler der neunten Jahrgangsstufe (Klieme et al., 2010; Köller et al., 2010). Wie in Abschnitt 4.2 dargestellt wird, überschneiden sich die Zielpopulationen beider Studien. Die Studien verwendeten für ihre Testung allerdings jeweils einen eigenen Lesetest.

Der vorliegende Artikel geht der Frage nach, welche Beziehung zwischen den in beiden Studien gemessenen Lesekompetenzkonstrukten besteht und inwieweit die auf den Testergebnissen der Studien basierenden Kompetenzverteilungen der Schülerinnen und Schüler in Deutschland empirisch vergleichbar sind.

Dargestellt werden zunächst die Zielsetzungen und die theoretischen Definitionen der Lesekompetenz in den beiden Studien sowie ihre jeweilige Operationalisierung in Testaufgaben und in kriterial beschreibbaren Kompetenzstufenmodellen. Anhand einer gemeinsamen Stichprobe von Schülerinnen und Schülern, die sowohl die Lesetestaufgaben von PISA 2009 als auch die der Bildungsstandards bearbeiteten, werden die empirischen Verteilungen auf den Kompetenzstufen verglichen und der latente Zusammenhang zwischen den Lesetestwerten analysiert.

2. Das Konzept der Lesekompetenz

Lesekompetenz wird als ein zentraler Aspekt von Grundbildung angesehen. Lesen zu können stellt nicht nur eine wesentliche Voraussetzung für den weiteren Kompetenzerwerb in der Schule dar, sondern ist grundlegend für die Weiterentwicklung eigenen Wissens und eigener Fähigkeiten sowie für die Teilhabe am kulturellen und gesellschaftlichen Leben auch außerhalb institutioneller Kontexte (Naumann, Artelt, Schneider & Stanat, 2010). Dabei ist Lesen kein passiver Prozess, sondern eine aktive Konstruktionsleistung des Individuums, das die im Text enthaltenen Inhalte mit seinem jeweils individuellen Vor- und Weltwissen in Verbindung bringen muss (Artelt, Stanat, Schneider & Schiefele, 2001).

Die Fähigkeit, verstehend zu lesen, umfasst neben kognitiven Prozessen auch motivationale und emotionale Aspekte (Groeben & Hurrelmann, 2004). Theoretische Konzeptionen beschreiben den Zusammenhang zwischen Lesekompetenz und Lesemotivation, und es kann argumentiert werden, dass eine aktive Lesepraxis eine Motivation zum Lesen voraussetzt (Artelt, Naumann & Schneider, 2010). Auch empirisch lassen sich Zusammenhänge zwischen Lesekompetenz und Leseinteresse sowie Lesemotivation aufzeigen (Guthrie & Wigfield, 2000; Artelt et al., 2010).

Entsprechend berücksichtigen die theoretischen Rahmenkonzeptionen von Large Scale Assessments motivationale Aspekte der Lesekompetenz (bspw. OECD, 2009a). Diese werden jedoch konzeptionell und methodisch nicht als immanenter Aspekt der Lesekompetenz aufgefasst, sondern als mit ihr zusammenhängende, separat erfassbare Aspekte. Sowohl die Testaufgaben als auch die Kompetenzstufenmodelle fokussieren daher auf das Textverstehen. Dieser Definition folgt auch der vorliegende Text.

Die Bedeutung, die der Lesekompetenz als universelles Kulturwerkzeug zukommt, ist auch in einer sich verändernden Medienlandschaft nicht geringer geworden. Im Umkehrschluss ist bei einer geringen Lesekompetenz von einem Chancennachteil nicht nur für das Individuum, sondern auch für die Gesellschaft insgesamt auszugehen.

Nicht zuletzt aus diesen Gründen ist die Erfassung von Lesekompetenz als ein wichtiger Aspekt grundlegender Bildung fester Bestandteil von Schulleistungsstudien sowohl im Grundschulalter als auch am Ende der Pflichtschulzeit. Zu letzteren zählen die Kompetenzerfassung bei Fünfzehnjährigen im internationalen Vergleich in PISA und die Kompetenzmessung auf Basis der Bildungsstandards in der neunten Jahrgangsstufe im nationalen Ländervergleich. Lesekompetenz wird dabei je nach Fokus der Studie und der zu testenden Altersgruppe unterschiedlich definiert und erfasst.

PISA und auch die Studien auf der Grundlage der Bildungsstandards arbeiten zur differenzierteren Beschreibung der Lesekompetenz mit dem Verfahren des *Standard-setting*, aus dem Kompetenzstufenmodelle resultieren. Kompetenzstufen dienen einer kriteriumsorientierten Beschreibung von Kompetenz, die über eine rein numerische Quantifizierung hinausgehen soll. Sie werden gebildet, indem die aus der Messung resultierende, kontinuierliche quantitative Kompetenzskala in qualitativ unterscheidbare Abschnitte unterteilt wird. Die so entstandenen Kompetenzstufen lassen sich inhaltlich beschreiben und es wird definiert, über welche Kompetenzen Personen mit einer bestimmten Testpunktzahl verfügen, d.h. welche Aufgabenanforderungen sie mit einer bestimmten Wahrscheinlichkeit bewältigen können. Die Bildung der Kompetenzstufen, also die Bestimmung von Schwellen zwischen verschiedenen Abschnitten der kontinuierlichen Skala, wird dabei von inhaltlichen und psychometrischen Vorgaben bestimmt (Hartig & Klieme, 2006; Hartig & Jude, 2007; Pant, Tiffin-Richards & Köller, 2010).

Im Folgenden werden die Konzepte der Lesekompetenz in PISA und den Bildungsstandards, die jeweiligen Verfahren der Testentwicklung sowie die Kompetenzstufenbildung in beiden Studien dargestellt und zusammenfassend verglichen.

2.1 Das Programme for International Student Assessment (PISA)

Hintergrund und Zielsetzung

PISA als international vergleichende Schulleistungsstudie wird von der Organisation für wirtschaftliche Zusammenarbeit und Entwicklung (OECD) initiiert und erfasst die Fähigkeiten von 15-jährigen Schülerinnen und Schülern in den Bereichen der Lesekompetenz, der mathematischen Kompetenz und der naturwissenschaftlichen Kompetenz in den OECD-Staaten sowie in OECD-Partnerstaaten. PISA wird seit dem Jahr 2000 im

Abstand von drei Jahren regelmäßig international durchgeführt. In Deutschland erfolgt die Beteiligung im Auftrag der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland (KMK).

PISA ermöglicht Aussagen darüber, inwieweit Schülerinnen und Schüler gegen Ende ihrer Pflichtschulzeit über grundlegende Kompetenzen verfügen. Die Ergebnisse geben außerdem „Auskunft über demografische Kontextindikatoren sowie Merkmale der häuslichen und schulischen Umwelt, die mit den Kompetenzen in Zusammenhang stehen können, und erlauben die Analyse von Veränderungen über die Zeit“ auf Ebene der verschiedenen Bildungssysteme (Jude & Klieme, 2010, S. 13).

Die regelmäßige Teilnahme Deutschlands an PISA ermöglicht die systematische Sammlung von Informationen über das deutsche Bildungssystem und die Analyse von Veränderungen (Trends) in den Leistungen von Schülerinnen und Schülern sowie den Kontextbedingungen schulischer Lehr-Lernprozesse. PISA ist damit Teil der im Jahr 2006 verabschiedeten Gesamtstrategie der KMK zum Bildungsmonitoring (KMK, 2006), in der die Teilnahme an international vergleichenden Schulleistungstudien eine zentrale Rolle spielt. Die Ergebnisse dieser internationalen Vergleiche ergänzen die Informationen, die aus der zentralen Überprüfung der Bildungsstandards gewonnen werden, und gehen in die gemeinsame Bildungsberichterstattung von Bund und Ländern ein (Autorengruppe Bildungsberichterstattung, 2012).

Studienkonzeption und Konstruktdefinition

PISA ist in regelmäßigen Erhebungszyklen aufgebaut: Im Abstand von jeweils drei Jahren werden drei Kompetenzbereiche getestet, wobei die Schwerpunkte zwischen den Erhebungen wechseln. Im Jahr 2000 war Lesekompetenz der Schwerpunkt, gefolgt von Mathematik im Jahr 2003 und den Naturwissenschaften im Jahr 2006. In PISA 2009 bildete die Lesekompetenz erneut den inhaltlichen Schwerpunkt.

Das Verständnis von Lesekompetenz in PISA basiert auf dem Literacy-Konzept und umfasst nach Definition der OECD „die Fähigkeit einer Person, geschriebene Texte zu verstehen, zu nutzen, über sie zu reflektieren und sich mit ihnen auseinanderzusetzen, um eigene Ziele zu erreichen, das eigene Wissen und Potenzial weiterzuentwickeln und am gesellschaftlichen Leben teilzunehmen“ (OECD, 2010, S. 24). PISA fokussiert auf die Kompetenzen der Jugendlichen zum Ende der Pflichtschulzeit, daher werden grundlegende Lesefertigkeiten, die zu einem früheren Zeitpunkt im Kompetenzerwerb stattfinden, nicht separat getestet.

Die Rahmenkonzeption von PISA differenziert in ihrer Definition von Lesekompetenz empirisch unterscheidbare Teildimensionen (OECD, 2009a). Diese beziehen sich zum einen auf die Anforderungen, die eine Leseaufgabe an eine Leserin oder einen Leser stellen kann: a) Informationen suchen und extrahieren (Informationsentnahme), b) textbezogen Informationen kombinieren und Interpretationen ableiten (Integration und Interpretation), und c) über einen Text reflektieren und diesen bewerten (Reflexion). Zusätzlich zur Berichtslegung über die Gesamtskala der Lesekompetenz werden für diese Teildimensionen Kompetenzskalen gebildet und die Kompetenzstände darauf

separat berichtet. Darüber hinaus unterscheidet PISA in seinem Aufgabenmaterial zwischen kontinuierlichen und nichtkontinuierlichen Texten und berichtet auch hierfür die von den Schülerinnen und Schülern erreichten Kompetenzverteilungen auf separaten Skalen. Neben der Gesamtskala der Lesekompetenz, die in dem vorliegenden Artikel verwendet wird, werden also auf Basis des Aufgabenmaterials insgesamt fünf Teildimensionen der Lesekompetenz unterschieden, die jeweils auf spezifische Aspekte der Lesekompetenz fokussieren.

Unterschiedliche Inhalte der Lesetexte, die sich aus verschiedenen Funktionen, Situationen und Kontexten des Lesens ergeben, werden nicht als empirisch trennbare Teilkompetenzen angesehen. Sie gehen jedoch inhaltlich in die Entwicklung des Testmaterials ein, um eine breite und wirklichkeitsnahe Gestaltung des Aufgabenmaterials sicherzustellen. Dazu gehören entsprechend der Lebenswelt der Fünfzehnjährigen persönliche Texte, öffentliche und bildungsbezogene Texte sowie Texte mit berufsbezogenen Inhalten.

Aufgabenentwicklung zur Messung der Lesekompetenz in PISA

Die Entwicklung von Testaufgaben für PISA basiert auf der Rahmenkonzeption der Lesekompetenz und erfolgt in Zusammenarbeit von mehreren internationalen Expertengremien. Hierzu gehören das internationale PISA-Konsortium, die internationalen Expertenausschüsse unter Federführung der OECD sowie nationale Sachverständige in allen Teilnehmerstaaten. Aus diesem Vorgehen resultieren Testaufgaben, die eine breite kulturelle Diversität aufweisen und in den Lesetexten die unterschiedlichen Lebenskontexte von Fünfzehnjährigen berücksichtigen, die an dieser international vergleichenden Studie teilnehmen (OECD, 2012). Diese Lesetexte und Testaufgaben werden von den Teilnehmerstaaten hinsichtlich ihrer kulturellen Angemessenheit eingeschätzt.

Die Testentwicklung erfolgt in der Regel englischsprachig, so dass die Aufgaben vor der Testung in die jeweilige Landessprache übersetzt werden müssen. In einem mehrstufigen Übersetzungsprozess wird sichergestellt, dass die Lesetexte in allen sprachlichen Fassungen authentisch wirken und die Aufgaben sowohl inhaltlich als auch in Bezug auf das sprachliche Anspruchsniveau der Vorlage entsprechen. Alle neu entwickelten Testaufgaben werden vor ihrem Einsatz in der Hauptuntersuchung in einer internationalen Pilotierungsstudie erprobt. Die Ergebnisse dieser Pilotierung liefern Hinweise auf die internationale Vergleichbarkeit der Testaufgaben, auf ihre Messeigenschaften in allen Teilnehmerstaaten sowie auf die empirische Abdeckung der Kompetenzstufen. Auf Basis dieser Ergebnisse werden die Testaufgaben für die Hauptstudie ausgewählt. In der 2009 durchgeführten Hauptstudie wurden neben neu entwickelten Testaufgaben auch bereits existierende Aufgaben aus PISA 2000 eingesetzt (OECD, 2010).

Im Erhebungsdesign von PISA werden innerhalb von zwei Zeitstunden Testaufgaben aus drei verschiedenen Kompetenzbereichen bearbeitet. Um mehr Testmaterial einsetzen zu können, als in der vorgesehenen Zeit bearbeitet werden kann, wird ein sogenanntes „Multi-Matrix-Design“ verwendet. Schülerinnen und Schüler bearbeiten

unterschiedliche Testhefte, die jeweils nur eine Teilmenge der insgesamt verwendeten Testaufgaben enthalten (Naumann et al., 2010).

Um trotz der eingeschränkten Testzeit eine reliable Messung zu ermöglichen, sind die verwendeten Lesetexte eher kurz, ihr Umfang beträgt selten mehr als eine Seite. In PISA 2009 wurden international 29 Lesetexte (Stimuli) mit insgesamt 131 Testaufgaben in der Hauptstudie eingesetzt. Pro Lesetext waren also durchschnittlich fünf Fragen zu beantworten (OECD, 2009b).

Etwa zwei Drittel der Testaufgaben in PISA beziehen sich auf kontinuierliche Lesetexte (60%), die restlichen Texte sind nicht kontinuierlich oder weisen ein gemischtes Format auf. Die Hälfte der Testaufgaben ist in Multiple-Choice-Antwortformaten zu bearbeiten (48%), bei den restlichen Aufgaben sind offene Antworten zu geben. Diese können entweder aus eindeutigen Richtiglösungen („closed constructed response“, 10%) oder aus kurzen („short response“, 8%) beziehungsweise längeren offenen Antworten bestehen („open constructed response“, 34%). Die Antworten der Schülerinnen und Schüler werden bei offenen Formaten durch geschulte Rater beurteilt.

Die oben dargestellten Teildimensionen der Lesekompetenz spiegeln sich wie folgt in den Testaufgaben wider: Die Hälfte der Aufgaben bezieht sich auf die Teildimension der Integration und Interpretation, jeweils ein Viertel der Aufgaben auf die Teildimensionen der Informationsentnahme beziehungsweise der Reflexion (vgl. Tabelle 1).

Teildimensionen Antwortformate	Informations- entnahme	Integration und Interpretation	Reflexion	Gesamt
Multiple choice	6	38	8	52 (40%)
Complex multiple choice	3	6	1	10 (8%)
Closed constructed response	9	4	0	13 (10%)
Short response	10	1	0	11 (8%)
Open constructed response	3	18	24	45 (34%)
	31 (24%)	67 (51%)	33 (25%)	131 (100%)

Tab. 1: Lesetestaufgaben in PISA 2009, Verteilung auf Teildimensionen und Antwortformate (vgl. OECD, 2012, S. 41)

Die PISA-Lesetexte beziehen sich zu etwa gleichen Teilen auf persönliche, öffentliche und bildungsbezogene Inhalte; berufsbezogene Inhalte sind demgegenüber etwas weniger stark repräsentiert (OECD, 2009b).

Das Kompetenzstufenmodell für den Bereich Lesen in PISA

PISA nutzt zur Beschreibung der Lesekompetenz eine kontinuierliche Skala. In der ersten PISA-Erhebung im Jahr 2000 wurde diese Skala auf einen Mittelwert von 500 und eine Standardabweichung von 100 für die damals teilnehmenden OECD-Staaten normiert. Im Jahr 2009 wurden die verwendeten Testaufgaben zunächst unabhängig kalibriert und anschließend durch ein *Equating* auf der Skala von PISA 2000 verankert (OECD, 2012). Durch die Veränderung der Zusammensetzung der OECD-Staaten sowie Veränderungen im Niveau der Lesekompetenz der 15-Jährigen seit dem Jahr 2000 liegt für PISA 2009 der Durchschnitt aller OECD-Staaten bei 493 Punkten, die Standardabweichung bei 93 Punkten (OECD, 2010).

Das Kompetenzstufenmodell für den Bereich Lesen von PISA 2000 umfasste fünf Kompetenzstufen (Adams & Wu, 2002). Für PISA 2009 wurde das Kompetenzstufenmodell analog zur entsprechend überarbeiteten Rahmenkonzeption erweitert, und zwar speziell im unteren und oberen Kompetenzbereich. Die unterste Kompetenzstufe I wurde aufgeteilt in die Stufen Ib und Ia, und die Kompetenzstufe VI wurde im oberen Kompetenzbereich neu hinzugefügt. Dadurch ist eine stärkere Differenzierung möglich als noch im Jahr 2000. Das Kompetenzstufenmodell für die Lesekompetenz in PISA 2009 umfasst jetzt die in Tabelle 2 dargestellten sieben kriterial beschreibbaren Kompetenzstufen (OECD, 2010).

Das Kompetenzstufenmodell von PISA 2009 definiert mit der Stufe Ib explizit eine untere Grenze, unterhalb derer sich die mit den PISA-Testaufgaben erfasste Lesekompetenz nicht mehr beschreiben lässt. Über die Lesekompetenz von Personen, die sich unterhalb dieser Kompetenzstufe befinden, kann anhand von PISA lediglich die Aussage gemacht werden, welche Leseanforderungen sie mit großer Wahrscheinlichkeit *nicht* lösen. Ihre Lesekompetenz kann anhand der PISA-Testaufgaben also nicht inhaltlich kriterial beschrieben werden.

Die Definition der Kompetenzstufen basiert auf inhaltlichen und psychometrischen Kriterien und erfolgte in einem zweistufigen Verfahren. Im ersten Schritt wurden die Testaufgaben in Relation zu den Kompetenzaspekten der Rahmenkonzeption analysiert, d.h. die einzelnen Testaufgaben wurden im Hinblick darauf beurteilt, inwieweit sie die Kompetenzaspekte der Rahmenkonzeption repräsentieren. Daraus resultierte die Beschreibung der kognitiven Anforderungen jeder Testaufgabe – von leichten Aufgaben bis hin zu schwierigen Aufgaben. Diese Beschreibungen wurden zusammen mit den empirischen Schwierigkeiten jeder Testaufgabe verwendet, um in einem zweiten Schritt die Grenzen zwischen den Kompetenzstufen zu definieren. Die Setzung von Schwellenwerten („cut-offs“) bestimmte letztlich die Zuordnung von Testaufgaben zu Kompetenzstufen (vgl. zum genauen Vorgehen OECD, 2012). Die resultierenden Kompetenzstufenmodelle bilden die Grundlage für die Ergebnisberichte von PISA.

Stufe	Wertebereich	Wozu die Schülerinnen und Schüler auf der jeweiligen Kompetenzstufe im Allgemeinen in der Lage sind
VI	> 698	Jugendliche auf dieser Stufe können Schlussfolgerungen, Vergleiche und Gegenüberstellungen detailgenau und präzise anstellen. Dabei entwickeln sie ein volles und detailliertes Verständnis eines oder mehrerer Texte und verbinden dabei unter Umständen gedanklich Informationen aus mehreren Texten miteinander.
V	626-697	Hierbei kann auch die Auseinandersetzung mit ungewohnten Ideen gefordert sein, genauso wie der kompetente Umgang mit konkurrierenden Informationen und abstrakten Interpretationskategorien sowie hohe Präzision im Umgang mit zum Teil unauffälligen Textdetails.
IV	553-625	Jugendliche auf dieser Stufe können sowohl mehrere tief eingebettete Informationen finden, ordnen und herausfinden, welche davon jeweils relevant sind, als auch ausgehend von Fachwissen eine kritische Beurteilung oder Hypothese anstellen. Die Aufgaben dieser Stufe setzen in der Regel ein volles und detailliertes Verständnis von Texten voraus, deren Inhalt oder Form ungewohnt ist. Zudem muss mit Konzepten umgegangen werden können, die im Gegensatz zum Erwarteten stehen.
III	480-552	Aufgaben dieser Kompetenzstufe erfordern vom Leser/von der Leserin, linguistischen oder thematischen Verknüpfungen in einem Text über mehrere Abschnitte zu folgen, oftmals ohne Verfügbarkeit eindeutiger Kennzeichen im Text, um eingebettete Informationen zu finden, zu interpretieren und zu bewerten oder um psychologische oder philosophische Bedeutungen zu erschließen. Insgesamt muss ein genaues Verständnis langer oder komplexer Texte, deren Inhalt oder Form ungewohnt sein kann, unter Beweis gestellt werden.
II	408-479	Aufgaben dieser Kompetenzstufe erfordern vom Leser/von der Leserin, vorhandenes Wissen über die Organisation und den Aufbau von Texten zu nutzen, implizite oder explizite logische Relationen (z.B. Ursache-Wirkungs-Beziehungen) über mehrere Sätze oder Textabschnitte zu erkennen, mit dem Ziel, Informationen im Text zu lokalisieren, zu interpretieren und zu bewerten. Einige Aufgaben verlangen vom Leser/von der Leserin, einen Zusammenhang zu begreifen oder die Bedeutung eines Wortes oder Satzes zu analysieren. Häufig sind die benötigten Informationen dabei nicht leicht sichtbar oder Passagen des Textes laufen eigenen Erwartungen zuwider.
Ia	335-407	Jugendliche auf dieser Stufe können innerhalb eines Textabschnitts logischen und linguistischen Verknüpfungen folgen, mit dem Ziel, Informationen im Text zu lokalisieren oder zu interpretieren; im Text oder über Textabschnitte verteilte Informationen aufeinander beziehen, um die Absicht des Autors zu erschließen. Bei Aufgaben dieser Stufe müssen unter Umständen auf der Grundlage eines einzigen Textbestands Vergleiche und Gegenüberstellungen vorgenommen werden oder es müssen, ausgehend von eigenen Erfahrungen oder Standpunkten, Vergleiche angestellt oder Zusammenhänge zwischen dem Text und nicht im Text enthaltenen Informationen erkannt werden.
Ib	262-334	Aufgaben dieser Kompetenzstufe erfordern vom Leser/von der Leserin, in einem Text zu einem vertrauten Thema eine oder mehrere unabhängige, explizit ausgedrückte Informationen zu lokalisieren, das Hauptthema oder die Absicht des Autors zu erkennen oder einen einfachen Zusammenhang zwischen den im Text enthaltenen Informationen und allgemeinem Alltagswissen herzustellen. Die erforderlichen Informationen sind in der Regel leicht sichtbar, und es sind nur wenige beziehungsweise keine konkurrierenden Informationen vorhanden. Der Leser wird explizit auf die entscheidenden Elemente in der Aufgabe und im Text hingewiesen.
unter Ib	< 262	Jugendliche auf dieser Stufe können in einem kurzen, syntaktisch einfachen Text aus einem gewohnten Kontext, dessen Form vertraut ist (z.B. in einer einfachen Liste oder Erzählung), eine einzige, explizit ausgedrückte Information lokalisieren, die leicht sichtbar ist. Der Text enthält in der Regel Hilfestellungen für den Leser, wie Wiederholungen, Bilder oder bekannte Symbole. Es gibt kaum konkurrierende Informationen. Bei anderen Aufgaben müssen einfache Zusammenhänge zwischen benachbarten Informationsteilen hergestellt werden.
		Keine kriteriale Beschreibung der Kompetenz möglich.

Tab. 2: Wertebereich auf der PISA-Skala und inhaltliche Beschreibung der Kompetenzstufen von PISA 2009 (Naumann et al., 2010)

2.2 Die Studien auf Basis der Bildungsstandards

Hintergrund und Zielsetzung

Vor dem Hintergrund des unbefriedigenden Abschneidens der Schülerinnen und Schüler in deutschen Schulen in internationalen Vergleichsstudien seit TIMSS (Baumert, Bos & Lehmann, 2000) wurde eine Reihe von Maßnahmen zur Qualitätssicherung und -entwicklung im deutschen Schulsystem beschlossen (KMK, 1997, 2006). Eine wichtige Grundlage dieser Maßnahmen bilden die abschlussbezogenen länderübergreifenden Bildungsstandards, die seit 2003/2004 unter anderem für den Mittleren Schulabschluss im Fach Deutsch vorliegen. Bildungsstandards stellen verbindlich und fachspezifisch formulierte Kernziele schulischer Bildung dar und benennen jene Kompetenzen, die Schülerinnen und Schüler im jeweiligen Fach bis zu einem bestimmten Zeitpunkt in ihrer Bildungsbiografie im Regelfall erworben haben sollten. Bildungsstandards sollen zudem helfen, den Fokus von einer Input- hin zu einer Output-Orientierung zu verlagern, bei der nicht Lerninhalte, sondern Bildungsergebnisse die zentrale Größe in der Steuerung des deutschen Bildungssystems darstellen.

Im Zuge der Einführung der Bildungsstandards wurde im Jahr 2004 das Institut zur Qualitätsentwicklung im Bildungswesen (IQB) als An-Institut der Humboldt-Universität zu Berlin gegründet, dessen Auftrag in der Weiterentwicklung, Operationalisierung, Normierung und Überprüfung des Erreichens der Bildungsstandards besteht. Mit den sog. Plöner Beschlüssen aus dem Jahr 2006 (KMK, 2006) wurde zudem entschieden, auf Basis der Bildungsstandards regelmäßig nationale Ländervergleiche durchzuführen. Diese nationalen Studien ergänzen seit dem Jahr 2009 die internationalen Schulleistungsvergleiche PISA (Sekundarstufe I) und IGLU/PIRLS sowie TIMSS (Primarstufe), indem sie detaillierte Informationen darüber bereitstellen, inwieweit die Schülerinnen und Schüler in den 16 deutschen Ländern die in den Bildungsstandards definierten Leistungserwartungen erfüllen (Köller et al., 2010; Stanat et al., 2012).

Studienkonzeption und Konstruktdefinition

Die Durchführung der Ländervergleiche auf Basis der Bildungsstandards im Bereich der Sekundarstufe I ist an die internationalen PISA-Studien angebunden und erfolgt daher ebenfalls im dreijährigen Zyklus. Dabei wird jeweils eine Auswahl der Kompetenzbereiche getestet, für die in der Sekundarstufe I Bildungsstandards vorliegen. Im Ländervergleich 2009 wurden Kompetenzen in der Verkehrssprache Deutsch, darunter die Lesekompetenz, sowie in der ersten Fremdsprache Englisch bzw. Französisch erhoben. In diesem Durchgang lag auch bei PISA der Schwerpunkt auf der Lesekompetenz. Während jedoch bei PISA die internationale Rahmenkonzeption die Grundlage des Lesetests bildete (vgl. Abschnitt 2.1, Studienkonzeption und Konstruktdefinition), basierte die Messung der Lesekompetenz im Ländervergleich auf den Bildungsstandards.

In den Bildungsstandards im Fach Deutsch für den Mittleren Schulabschluss (MSA) umfasst der Kompetenzbereich „Lesen – mit Texten und Medien umgehen“ zum einen das Verstehen von literarischen Texten und Sachtexten, zum anderen jedoch auch die

ästhetische Dimension des Lesens sowie die Förderung von Leseinteresse und Lese Freude (KMK, 2004a). Da die beiden letztgenannten Aspekte im Kontext eines Large Scale Assessments kaum mit Testaufgaben operationalisierbar sind, liegt der Fokus der Konstruktdefinition auf den Standards „literarische Texte verstehen und nutzen“ sowie „Sach- und Gebrauchstexte verstehen und nutzen“ (vgl. KMK, 2004a, S. 14-15). Gemäß dieser Definition wird Lesekompetenz in den Studien des IQB auf Basis der Bildungsstandards als die Fähigkeit verstanden, auch längere Texte sowie nichtlineare Texte, wie z.B. Schaubilder, zu verstehen und im Detail zu erfassen. Hierzu gehören die zielgerichtete Entnahme und das Ordnen von Informationen aus einem Text sowie die Verknüpfung von verschiedenen im Text enthaltenen Informationen miteinander und mit dem eigenen Vorwissen. Schülerinnen und Schüler, die die Regelstandards erreichen, sind ferner in der Lage, diese Informationen zu interpretieren und hinsichtlich ihrer Aussagekraft und Wertung zu prüfen, um die Intention eines Textes zu erkennen und begründete Schlussfolgerungen zu ziehen (vgl. Böhme, Neumann & Schpolowski, 2010). Lesekompetenz umfasst somit Rezeptionsleistungen vom Detailverstehen über Globalverstehen bis hin zur kritischen Reflexion von Texten. Des Weiteren beinhaltet die Definition der Lesekompetenz im Rahmen der Bildungsstandards auch die Kenntnis sprachlicher Gestaltungsmittel sowie das Erkennen und Unterscheiden verschiedener literarischer Gattungen, Textsorten und -funktionen, die in den Bildungsstandards als weitere wesentliche Aspekte des Verstehens und der Nutzung von Texten genannt werden.

Auf Basis dieser Konstruktdefinition erfolgt die Entwicklung von Testaufgaben zur Erfassung der Lesekompetenz in den nationalen Ländervergleichsstudien.

Entwicklung von Testaufgaben zur Messung der Lesekompetenz

Bei der Entwicklung von Testaufgaben steht der Innovationsanspruch der Bildungsstandards im Vordergrund, der – in Anlehnung an PISA – unter anderem in der Kompetenzorientierung sowie in der stärkeren Berücksichtigung kontinuierlicher und diskontinuierlicher Sachtexte zum Ausdruck kommt. Gleichzeitig wird ein gewisses Maß an curricularer Validität angestrebt, die durch die Genese der Bildungsstandards gewährleistet wird. Diese beruhen auf fachspezifischen Kompetenzmodellen, die sowohl international anerkannte Modelle wie die theoretischen Grundlagen der PISA-Studien berücksichtigen als auch „aus der Erfahrung der Schulpraxis heraus entwickelt wurden“ (KMK, 2004a, S. 3). Zudem soll der im Folgenden beschriebene Aufgabenentwicklungsprozess durch die Einbeziehung unterrichtender Lehrkräfte sicherstellen, dass Elemente der aktuellen unterrichtlichen Gestaltung in die Testentwicklung einfließen.

Die Testaufgaben für die bildungsstandardbasierten Kompetenzmessungen werden von erfahrenen Lehrkräften aus ganz Deutschland entwickelt. Diese werden durch Expertinnen und Experten aus der jeweiligen Fachdidaktik, der empirischen Bildungsforschung und Psychometrie geschult und beraten. Jede Aufgabe wird in einem mehrstufigen Review-Prozess wiederholt beurteilt, in kleinen Schülergruppen präpilotiert, überarbeitet und schließlich im Rahmen einer größeren Pilotierungsstudie erprobt. Basierend auf den empirischen Ergebnissen dieser Pilotierung werden besonders geeig-

nete Aufgaben ausgewählt, gegebenenfalls erneut überarbeitet und in einer für ganz Deutschland repräsentativen Erhebung normiert (Normierungsstudie). Die aus diesem Entwicklungsprozess resultierenden Aufgaben werden schließlich im Rahmen verschiedener Studien zur Kompetenzmessung auf Basis der Bildungsstandards eingesetzt (Pöhlmann, Neumann, Tesch & Köller, 2010; Granzer, 2009).

Im 2009 durchgeführten Ländervergleich Sprachen kamen für das Fach Deutsch im Bereich Lesen acht authentische Lesetexte (Stimuli) zum Einsatz, darunter drei literarische Texte und fünf Sachtexte. Die Länge der Textstimuli betrug zwischen einer halben und drei Seiten, wobei etwa anderthalb Seiten die Regel waren. Es handelte sich in jedem Fall um Fließtexte, die jedoch teilweise durch Abbildungen ergänzt oder aus mehreren Quellen zusammengeführt wurden. Pro Stimulus waren zwischen fünf und zwölf Testaufgaben zu bearbeiten, die teilweise aus mehreren Teilaufgaben (Items) bestanden. Insgesamt kamen 75 Testaufgaben zum Einsatz, zu denen insgesamt 99 Items bearbeitet werden sollten. Aufgrund der begrenzten Testzeit wurde – wie bei der PISA-Erhebung – ein komplexes Testdesign verwendet (vgl. Böhme, Leucht et al., 2010), bei dem jedem Schüler bzw. jeder Schülerin nur ein Teil der gesamten Aufgabenmenge vorgelegt wurde.

Die Antwortformate umfassten neben Multiple-Choice-Aufgaben mit vier Antwortalternativen (ca. 44% der Teilaufgaben) auch sog. Richtig-Falsch-Aufgaben, bei denen für vorgegebene Aussagen entschieden werden musste, ob sie zutreffen oder nicht (12%). Ein weiteres geschlossenes Antwortformat stellten Zuordnungs- und Umordnungsaufgaben dar (8%). Bei offenen Antwortformaten wurde unterschieden zwischen Kurzantworten, für die eine intersubjektiv eindeutige Richtiglösung definiert werden kann (15%), und Aufgabenstellungen, deren Beantwortung überwiegend Interpretationsleistungen erfordert und bei deren Auswertung daher ein gewisser Beurteilungsspielraum besteht (21%).

Das Kompetenzstufenmodell für den Bereich Lesen

Seit April 2009 liegt für den Bereich „Lesen – mit Texten und Medien umgehen“ ein fachdidaktisch fundiertes Kompetenzstufenmodell zu den Bildungsstandards für den Mittleren Schulabschluss (MSA) vor (Bremerich-Vos et al., 2010). Dieses Modell wurde für Schülerinnen und Schüler am Ende der Pflichtschulzeit entwickelt, die den MSA anstreben. Die Datenbasis des Modells bildet eine Normierungsstudie des IQB aus dem Jahr 2008, in die Schülerinnen und Schüler aus den Jahrgangsstufen acht bis zehn aller Schularten des allgemeinbildenden deutschen Schulsystems einbezogen wurden. Die Festlegung der zugrunde liegenden Metrik orientierte sich am internationalen Kompetenzstufenmodell von PISA aus dem Jahr 2000 (Bremerich-Vos et al., 2010); d.h. die Kompetenzskala wurde so definiert, dass sie in der Population der Neuntklässler in Deutschland einen Mittelwert von 496 Punkten und eine Standardabweichung von 92 Punkten aufweist.

Wie in anderen Modellen zu den Bildungsstandards wurden fünf Kompetenzstufen definiert, deren Schwellen im gleichen Abstand zueinander liegen. Für den Bereich der

Lesekompetenz in Deutsch beträgt die Stufenbreite jeweils 80 Punkte. Die unterste sowie die oberste Stufe sind nach unten bzw. nach oben offen. Die Fähigkeit jeder Person mit einem gültigen Testwert kann also einer kriterial beschreibbaren Kompetenzstufe zugeordnet werden.

Die zur Differenzierung der Kompetenzstufen verwendeten Cut-off-Werte und die inhaltlichen Beschreibungen der Kompetenzstufen sind in Tabelle 3 dargestellt.

Da die auf den einzelnen Stufen beschreibbaren Aspekte von Lesekompetenz mit den in den Bildungsstandards formulierten Kompetenzerwartungen verknüpft wurden, geben diese Hinweise darauf, inwieweit Regelanforderungen erfüllt oder verfehlt werden. Leseleistungen auf Kompetenzstufe III werden als das Erreichen der Regelstandards für den MSA interpretiert. Weiterhin wurde festgelegt, dass die Kompetenzstufe II als Mindeststandards zu interpretieren ist, der von allen Schülerinnen und Schülern, die den MSA anstreben, erreicht werden sollte.

Die Stufenbeschreibungen beziehen sich auf die in der Konstruktdefinition genannten Verständnisleistungen wie das Lokalisieren von Informationen, deren Verknüpfung zur Ableitung von Schlussfolgerungen sowie Interpretation und Reflexion (vgl. Abschnitt 2.2, Studienkonzeption und Konstruktdefinition). Ferner wird explizit auf unterschiedliche Textsorten, Textaufbau, Erzählperspektive sowie die Rolle des Antwortformats eingegangen. Die teilweise gegebene Aufgabennähe der Deskriptoren hängt damit zusammen, dass das Modell nicht allein für die Zwecke des Bildungsmonitoring auf Systemebene erstellt wurde, sondern auch im Rahmen von Unterrichtsentwicklung Anwendung finden soll. In diesem Zusammenhang kommt den exemplarischen Aufgabenstellungen, die eine Konkretisierung von Anforderungen ermöglichen, eine größere Bedeutung zu.

2.3 Zusammenfassender Vergleich

In den vorangegangenen Abschnitten wurden die Definitionen der Lesekompetenz in PISA und der Bildungsstandards dargestellt und beschrieben, wie diese Konstrukte mit Testaufgaben und Kompetenzstufenmodellen operationalisiert werden. Diese Informationen werden nun im Folgenden vergleichend gegenübergestellt.

Hintergrund und Zielsetzung der Studien

Beide hier dargestellten Studien erfassen die Lesekompetenz von Schülerinnen und Schülern am Ende der Sekundarstufe I. PISA wählt dabei eine altersbasierte Populationsdefinition der Fünfzehnjährigen, um einen vom jeweiligen Schulsystem unabhängigen, internationalen Vergleich der Kompetenzen innerhalb einer Alterskohorte zu ermöglichen. Den Ländervergleichsstudien auf Basis der Bildungsstandards liegt dagegen eine jahgangsbasierte Populationsdefinition zugrunde, d.h. es werden Schülerinnen und Schüler in der neunten Jahrgangsstufe getestet. Diese beiden Zielgruppen überschneiden sich in der Schülerpopulation in Deutschland zu ca. 50% (Jude & Klieme,

Stufe	Wertebereich	Wozu die Schülerinnen und Schüler auf der jeweiligen Kompetenzstufe im Allgemeinen in der Lage sind
V	> 640	<p>Schülerinnen und Schülern lösen hier nicht nur mit hoher Sicherheit Aufgaben der Stufen I bis IV. Es gelingt ihnen auch, Interpretationshypothesen plausibel zu beurteilen und in einem argumentativen Text zentrale Thesen zu identifizieren. Dabei geht es auch um das Erkennen der argumentativen Funktion von Beispielen. Ebenso können in einem umfangreichen Text einzelne Stellen als Fälle von positiven oder negativen Bewertungen identifiziert werden. Wörter und Wortgruppen in fiktionalen Texten, die für die Interpretation zentral sind, können verständlich erläutert werden. Bei einem weiteren Typ von (Teil-) Aufgaben, die auf dieser Stufe gelöst werden können, ist eine nur partiell ausgeführte schematische Darstellung relevanter Elemente eines langen Sachtexts vorgegeben. Fehlende Elemente sind einzufügen, was nur gelingt, wenn im Text verstreute Informationen identifiziert und verknüpft werden können. Darüber hinaus müssen für die korrekte Platzierung der Aufgaben auf Kompetenzstufe V verlangt, selbstständig einen mehr oder weniger umfangreichen Text zu produzieren. Insgesamt werden auf der Stufe V Leistungen gezeigt, die vermutlich nur unter sehr günstigen schulischen und außerschulischen Lerngelegenheiten erreichbar sind. Wir interpretieren diese Stufe dementsprechend im Sinne der Erreichung eines Maximalstandards.</p>
IV	560-639	<p>Auf dieser Stufe können Aufgaben der bisher beschriebenen Arten gelöst werden, das heißt unter anderem Aufgaben zur Strukturierung von Texten in Abschnitte und zur Angabe der Textsorte. Neben Aufgaben, die sich auf die Verknüpfung von Informationen über Abschnitte hinweg oder auch auf lokale Relationen von Informationen beziehen, sind auch noch Fälle von Lokalisierungsaufgaben im Spiel. Hier gibt es allerdings anders als auf den unteren Stufen kaum noch eine Identität oder offensichtliche Ähnlichkeit von Aufgabenformulierung und Darstellung im Text. Unter Berücksichtigung des gesamten (narrativen) Textes können darüber hinaus verschiedenen Figuren Merkmale zugeordnet werden. Schülerinnen und Schüler können auch über einen Text verstreute Varianten der Bezeichnung eines Gegenstandes nennen und sich begründet für eine von mehreren Interpretationen eines Textabschnitts entscheiden. Im Text nicht explizit genannte, sondern zu erschließende Wissensbestände von Figuren werden ebenso erkannt wie das Motiv eines Erzählers, einen Sachverhalt auf eine bestimmte Weise darzustellen. Auf dieser Stufe können auch Aufgaben gelöst werden, die sich auf Kombinationen von Grafik und kontinuierlichem Text beziehen. Dabei müssen unter anderem Informationen in beiden Formaten miteinander abgeglichen werden und es geht auch darum, den partiell hierarchischen Charakter der Grafik, das heißt die Einbettung einer Liste in eine andere Liste, zu verstehen. Auf Stufe IV werden somit insgesamt Leistungen gezeigt, die bereits über den Vorgaben der KMK liegen, so dass wir diese Stufe im Sinne der Erreichung eines Regelstandards plus interpretieren.</p>
III	480-559	<p>Über die bereits auf den Stufen I und II lösbaren Aufgaben hinaus werden hier Aufgaben gemeistert, in deren Rahmen unter anderem bei längeren Texten vorgegebene Absatzüberschriften in die richtige Reihenfolge zu bringen sind. Bei einem anderen auf die Textgliederung zielenden Aufgabentyp sind Absatzüberschriften Zeilenintervalle zuzuordnen. Hinzu kommen Aufgaben, bei denen bei komplexeren Texten aus mehreren vorgegebenen Alternativen das zutreffende Textthema zu wählen ist. Es gelingt auch – ebenfalls im Blick auf den ganzen Text – im Kontext einer Mehrfachwahlaufgabe die Erzählperspektive anzugeben. Weitere Aufgaben zielen auf die Differenzierung von Textsorten, zum Beispiel auf die Differenz von historisch und phantastisch, und darauf anzugeben, welches von mehreren Merkmalen am ehesten geeignet ist, einen Text der Phantastik zuzuweisen. Es werden auch Aufgaben gelöst, bei denen die Bedeutung von Wörtern kontextuell erschlossen werden muss. Bereits auf ein zentrales Element einer Interpretation als Resultat eines komplexen Schlussprozesses zielt die auf diesem Niveau beantwortbare Frage, auf welche von vier genannten Gruppen sich ein Personalpronomen bezieht. Schülerinnen und Schüler, die diese oder eine höhere Stufe erreichen, zeigen Leistungen, wie sie in den Bildungs-</p>

Stufe	Wertebereich	Wozu die Schülerinnen und Schüler auf der jeweiligen Kompetenzstufe im Allgemeinen in der Lage sind
II	400-479	<p>standards der KMK für den Bereich Lesen verfasst sind. Die Stufe III kann dementsprechend im Sinne der Erreichung von Regelstandards interpretiert werden.</p> <p>Auch auf dieser Stufe werden vor allem Aufgaben gelöst, bei denen es um Lokalisieren geht. Die gesuchten Informationen sind häufig weniger auffällig platziert. Über Aufgaben hinaus, bei denen Relationen von im Text benachbarten Informationen zu explizieren sind, werden auch einige Aufgaben bewältigt, bei denen die Verbindung von Informationen verlangt ist, die über mehrere Abschnitte verteilt sind. Vereinzelt werden hier auch Aufgaben gelöst, bei denen es darum geht, auf der Basis mehrerer Informationen eine angemessene Bezeichnung für das Handlungsmotiv eines Protagonisten auszuwählen. Häufig sind wie auf Stufe I Teile der Aufgabe mit Teilen der gesuchten Information bezeichnungs- oder bedeutungsgleich oder sie sind ihnen ähnlich. Das für die Lösung einiger Aufgaben nötige Weltwissen kann als weithin geteilt angesehen werden. So muss zum Beispiel im Kontext einer Aufgabe zu einem ICE-Prospekt geschlossen werden, dass man ein Problem mit einer Fahrkarte im Wagen mit einem Service Point zur Sprache bringen kann. Bei einer Reihe von Aufgaben, bei denen Lokalisieren gefragt ist, gibt es einen, manchmal auch mehrere attraktive Distraktoren im Text. Gelöst werden auch Aufgaben, bei denen ein mehrfaches Durchsuchen von Texten nach einzelnen Informationen gefordert ist. Neben Mehrfachwahlaufgaben kommen Aufgaben vor, bei denen ein Kurztext zu schreiben ist. Darüber hinaus gibt es einige Zuordnungs- und mehrteilige Wahrfalsch-Aufgaben. Schülerinnen und Schüler, die diese Kompetenzstufe erreichen, zeigen Leistungen, mit denen die in den Bildungsstandards der KMK implizierten Kompetenzstände noch nicht erreicht werden. Da einfache Verknüpfungen aber bereits gelingen und auch über den Text verstreute Informationen integriert werden können, interpretieren wir diese Stufe im Sinne der Erreichung von Mindeststandards.</p>
I	< 400	<p>Es können einzelne Informationen im Text lokalisiert werden, vor allem dann, wenn sie auffällig platziert sind (z. B. grafisch hervorgehoben oder zu Beginn oder am Ende von Absätzen). Gesucht werden häufig Bezeichnungen für Konkretes wie Eigennamen. Dabei kommt es gelegentlich vor, dass der Text plausible Optionen enthält, also Informationen, die der gesuchten mehr oder weniger ähneln. Es gelingt ebenfalls, auch längere und komplexere Texte auf das Vorhandensein einer einzelnen Information hin zu durchsuchen. Darüber hinaus werden Aufgaben gelöst, bei denen es um die Herstellung lokaler Kohärenz geht. Im Text benachbarte Informationen, deren Relationen nicht explizit sind, werden auf der Basis von verbreitetem Weltwissen beziehungsweise vor allem aufgrund von Sprachwissen miteinander verknüpft. Dabei ist das, was in der Aufgabe als Information bereits gegeben ist, oft mit dem Wortlaut beziehungsweise der Bedeutung eines Teils der Information, die im Text identifiziert werden muss, identisch oder verwandt. So ist zum Beispiel gefragt „Wer verfolgt die Prinzessin?“, und im Text ist der (Teil-) Satz „... dass sich die Prinzessin auf der Flucht vor den ... Trolen weigert weiterzulaufen“, zu ergänzen. Gelöst werden auch Aufgaben, bei denen angesichts eines strukturell einfachen und kurzen Textes das Textthema identifiziert werden muss beziehungsweise bei denen thematisch Relevantes von weniger Relevantem unterschieden werden muss. Es handelt sich meistens um geschlossene Aufgabenformate, vor allem um Mehrfachwahlaufgaben. Schülerinnen und Schüler, die auf dieser Stufe liegen, verfehlen deutlich die Vorgaben, wie sie in den Bildungsstandards der KMK für den Bereich Lesen verfasst sind. Die Stufe I beschreibt dementsprechend einen Leistungsabschnitt, auf dem auch ein Bildungsminimum im Sinne von Mindeststandards nicht erreicht wird.</p>

Tab. 3: Wertebereich auf der Skala der Bildungsstandards im Bereich Lesen in der Sekundarstufe I und inhaltliche Beschreibung der Kompetenzstufen des Lesemodells auf Basis der Bildungsstandards für den Mittleren Schulabschluss (Bremerich-Vos et al., 2010, S. 37)

2010). Es ist daher davon auszugehen, dass sich die im Durchschnitt erreichten Kompetenzniveaus dieser Populationen unterscheiden, weil die Fünfzehnjährigen in Deutschland in Abhängigkeit vom Einschulungsalter und der Bildungslaufbahn die Klassen 7 bis 10 besuchen (vgl. auch Mildner, Hochweber & Frey in diesem Heft).

Konstruktdefinition

Beide Studien greifen auf umfassende Modelle der Lesekompetenz zurück, die sowohl kognitive als auch motivationale Aspekte beinhalten. Letztere werden in beiden Studien bewusst nicht als Teil der über Testaufgaben zu messenden Kompetenz operationalisiert, sondern über begleitende Kontextfragebögen erhoben. Unterschiede finden sich in der konzeptionellen Ausrichtung der Konstruktdefinitionen jedoch dahingehend, dass PISA explizit von einem allgemeingültigen Literacy-Konzept ausgeht, das für alle teilnehmenden Staaten relevant ist, während das Konzept der Lesekompetenz bei den Bildungsstandards auf das deutsche Bildungssystem ausgerichtet ist. So flossen bereits in die Entwicklung der Bildungsstandards Erfahrungen aus der Schulpraxis in Deutschland ein. Darüber hinaus wird der Unterrichtsbezug jedoch in erster Linie durch den Aufgabenentwicklungsprozess realisiert, d.h. auf der Ebene der Operationalisierung.

PISA und die Studie auf Basis der Bildungsstandards differenzieren das Konstrukt der Lesekompetenz konzeptuell weiter aus, etwa durch die Unterscheidung von Textformaten. Beide Studien gehen also von einer Mehrdimensionalität der Lesekompetenz aus. In der Skalierung und Berichtlegung weist jedoch nur PISA empirisch fünf trennbare Teildimensionen der Lesekompetenz aus. Gemeinsam ist beiden Studien wiederum, dass sie die Lesekompetenz von Schülerinnen und Schülern (auch) auf einer Gesamtskala berichten.

Aufgabenentwicklung

Basierend auf der jeweiligen Konstruktdefinition erfolgt die Aufgabenentwicklung in beiden Studien durch geschulte Expertengruppen. Für die Studie auf Basis der Bildungsstandards sind dies Lehrerinnen und Lehrer aus Deutschland, die durch Fachdidaktiker und Psychometriker aus dem deutschsprachigen Raum angeleitet werden. Die Testaufgaben sollen einerseits kompetenzorientiert sein, in ihrer Gestaltung jedoch auch die Unterrichtspraxis in deutschen Schulen berücksichtigen und einen gewissen curricularen Bezug aufweisen. Bei PISA werden die Testaufgaben hingegen gemäß der internationalen Ausrichtung des Programms unter Beteiligung der Teilnehmerstaaten von internationalen Expertengruppen entwickelt. Sie beinhalten kulturell vielfältiges Lesematerial aus verschiedenen Staaten, das der Alltagswelt von Fünfzehnjährigen entsprechen soll. Die Testaufgaben von PISA sind nicht explizit curricular ausgerichtet.

Weitere Unterschiede zwischen beiden Studien bestehen dahingehend, dass für PISA die konzeptionell angenommenen Teildimensionen der Lesekompetenz nicht nur bei der Aufgabenentwicklung, sondern auch bei der Auswertung und Berichtlegung berücksichtigt werden. Darüber hinaus werden insgesamt mehr offene Antwortformate ver-

wendet als bei der Studie auf Basis der Bildungsstandards. Die in PISA verwendeten Lesetexte sind demgegenüber im Mittel kürzer als die Lesetexte im bildungsstandardbasierten Ländervergleich.

Kompetenzstufenmodell

PISA und die Studie auf Basis der Bildungsstandards veranschaulichen die kontinuierliche Kompetenzskala mit Hilfe von Kompetenzstufenmodellen. Das Kompetenzstufenmodell der Bildungsstandards für den Mittleren Schulabschluss im Bereich Lesen orientiert sich mit der Unterscheidung von fünf Kompetenzstufen am Modell für den Bereich Lesen von PISA 2000. Auch die Skalierung des Lesetests entspricht in wesentlichen Aspekten dem methodischen Vorgehen von PISA 2000.

Trotz ihrer Ähnlichkeit unterscheiden sich die Kompetenzstufenmodelle jedoch in zwei wesentlichen Punkten: Während das Kompetenzstufenmodell von PISA eine untere Grenze ausweist („unter Kompetenzstufe Ib“), unterhalb derer der Test definitionsgemäß nicht mehr in der Lage ist, valide zu messen, weicht das Modell der Lesekompetenz der Bildungsstandards hiervon ab: Die Kompetenzstufe I ist nach unten offen. Jeder Schülerin und jedem Schüler, die bzw. der an der Testung auf Basis der Bildungsstandards teilnimmt, wird also mindestens eine Lesekompetenz der Stufe I attestiert, sofern die Testung nicht vollständig verweigert wurde. In PISA hingegen wird die Lesekompetenz der Schülerinnen und Schüler unter Kompetenzstufe Ib nicht mehr kriterial beschrieben.

Ein weiterer wesentlicher Unterschied besteht in der Zielpopulation der Modelle. Während im PISA-Modell die Kompetenz der Fünfzehnjährigen aller Schulformate abgebildet werden soll, wurde das Modell der Bildungsstandards für die leistungstärkere Population derjenigen Schülerinnen und Schüler entwickelt, die in Deutschland den Mittleren Schulabschluss anstreben. Um in der leistungsheterogenen Population der Fünfzehnjährigen im oberen und unteren Bereich besser differenzieren zu können, weist das für PISA 2009 überarbeitete Modell der Lesekompetenz nun sieben Kompetenzstufen auf, d.h. es werden mehr Stufen ausgewiesen als im Modell der Bildungsstandards.

Die hier vergleichend gegenübergestellten konzeptuellen Gemeinsamkeiten und Unterschiede der beiden Studien im vorliegenden Artikel bilden den Hintergrund, vor dem die empirischen Ergebnisse im Folgenden diskutiert und interpretiert werden sollen. Ein inhaltlicher Vergleich der Kompetenzstufenbeschreibungen in den beiden Studien ist explizit nicht vorgesehen. Hierzu wäre eine umfassende Einschätzung durch eine Gruppe unabhängiger Experten nötig. Der Vergleich solcher Einschätzungen wäre zwar durchaus interessant, für die Fragestellung des vorliegenden Artikels, nämlich der empirischen Vergleichbarkeit der Ergebnisse beider Studien, jedoch nicht erforderlich.

Um die empirischen Befundmuster der Analysen dieses Artikels einordnen zu können, werden im Folgenden zunächst empirische Befunde vergleichbarer Studien dargestellt.

3. Empirische Befunde zur Vergleichbarkeit von internationalen und nationalen Kompetenztests

Die verschiedenen bisher in Deutschland durchgeführten nationalen und internationalen Schulleistungsstudien untersuchen teilweise den gleichen Kompetenzbereich, es werden jedoch unterschiedliche, speziell für den Zweck der jeweiligen Studie entwickelte Testaufgaben eingesetzt, die jeweils auf separaten Kompetenzdefinitionen basieren. Dabei stellt sich die Frage, inwieweit die Ergebnisse dieser unterschiedlichen Studien zum gleichen Kompetenzbereich vergleichbar sind. Im Folgenden wird dargestellt, wie sich diese Frage untersuchen lässt und welche Ergebnisse Analysen von Daten früherer Studien in Deutschland erbracht haben.

Verschiedene konzeptionelle und methodische Ansätze sind denkbar, um dieser Frage nachzugehen. So lässt sich analysieren, a) ob verschiedenen Studien die gleiche Kompetenzdefinition zugrunde liegt, b) ob die gleichen oder zumindest inhaltlich vergleichbare Testinstrumente verwendet werden und c) wie eng die aus unterschiedlichen Tests resultierenden Testwerte zusammenhängen. Darüber hinausgehend kann d) im Sinne eines *Equating* untersucht werden, ob Schülerinnen und Schüler auf Basis ihrer Testwerte in verschiedenen Tests der gleichen Kompetenz(stufe) zugewiesen werden.

Inzwischen haben verschiedene Publikationen die empirischen Zusammenhänge zwischen Kompetenztestwerten aus in Deutschland durchgeführten Schulleistungsstudien analysiert. Dabei sind zum einen die Zusammenhänge zwischen inhaltlich sehr ähnlichen Kompetenzkonstrukten, etwa der Lesekompetenz in verschiedenen Erhebungen, interessant. Zum anderen sind die Zusammenhänge zwischen unterschiedlichen Kompetenzbereichen bedeutsam, wie beispielsweise zwischen der Lesekompetenz und der mathematischen Kompetenz. Die Stärke dieser Zusammenhänge kann dann vergleichend interpretiert werden. Aus theoretischer Sicht wäre zu erwarten, dass die Zusammenhänge zwischen inhaltlich ähnlichen Kompetenzkonstrukten (Lesekompetenz in PISA und in Studien zu den Bildungsstandards) stärker ausfallen sollten als zwischen inhaltlich unterschiedlichen Kompetenzkonstrukten. Die Stärke der korrelativen Zusammenhänge zwischen verschiedenen Kompetenzbereichen kann also als Grundlage für die Interpretation von Korrelationen zwischen Messungen zum gleichen Kompetenzbereich in verschiedenen Studien dienen. Im Folgenden werden aktuelle empirische Ergebnisse zum Zusammenhang zwischen Tests zu unterschiedlichen Kompetenzbereichen sowie zwischen Tests zu denselben Kompetenzbereichen aus nationalen und internationalen Studien dargestellt.

Die latenten, also messfehlerbereinigten Korrelationen der Testwerte für die verschiedenen bei PISA erfassten Kompetenzbereiche liegen international im Durchschnitt zwischen $r = .77$ und $r = .89$ (PISA 2003, Leutner, Klieme, Meyer & Wirth, 2004, S. 167) beziehungsweise zwischen $r = .79$ und $r = .88$ (PISA 2006, vgl. OECD, 2009c, S. 215). Beispielsweise betragen die Korrelationen zwischen den Testwerten in Naturwissenschaften und Lesen in der deutschen Stichprobe von PISA 2000 $r = .87$, zwischen Naturwissenschaften und Mathematik $r = .83$, und zwischen Mathematik und Lesekompetenz $r = .84$ (vgl. Prenzel, Rost, Senkbeil, Häußler & Klopp, 2001, S. 221).

Bereits in PISA 2000 wurden die in Deutschland durchgeführten internationalen Schulleistungsstudien durch nationale Testverfahren ergänzt, so beispielsweise in den Bereichen Mathematik und Naturwissenschaften. Die nationalen Tests waren stärker an den Lehrplänen für Schulen in Deutschland orientiert als die internationalen Tests. Zwischen den internationalen und den national ergänzten Mathematiktests ergaben sich dabei Korrelationen von $r = .91$ (PISA 2000, vgl. Klieme, Neubrand & Lüdtke, 2001, S. 157) beziehungsweise von $r = .92$ (PISA 2003, vgl. Blum et al., 2004, S. 63). Die Korrelation zwischen den internationalen und den national ergänzten Tests im Bereich Naturwissenschaften betrug in PISA 2000 $r = .84$ (vgl. Prenzel et al., 2001, S. 224). Es kann festgestellt werden, dass die Zusammenhänge sich zwischen den Kompetenzbereichen (Mathematik und Naturwissenschaften) durchaus unterscheiden.

Seit Einführung der länderübergreifenden Bildungsstandards in Deutschland stellt sich zudem die Frage, wie die auf dieser Basis entwickelten Testaufgaben mit den internationalen Tests zusammenhängen. Für den Bereich der Mathematik untersuchten Hartig und Frey (2012) diese Fragestellung anhand von Daten aus der gemeinsamen Erhebung von PISA und Tests zu den Bildungsstandards in Mathematik im Jahr 2006. Der latente Zusammenhang zwischen den Testwerten im Bildungsstandards-Mathematiktest und im PISA-Mathematiktest fiel mit einer Korrelation von $r = .95$ signifikant stärker aus als die Korrelation zwischen Bildungsstandards-Mathematik und PISA-Lesen ($r = .77$) sowie zwischen Bildungsstandards-Mathematik und PISA-Naturwissenschaften ($r = .87$). Trotzdem kommen die Autoren auf Basis ihrer Analysen zu dem Schluss, dass ein differenziertes Modell, welches den Bildungsstandards-Mathematiktest und den PISA-Mathematiktest als zwei korrelierte, jedoch distinkte Dimensionen ausweist, die Varianz in den vorliegenden Daten besser erklären kann als ein eindimensionales Modell. Sie begründen dieses Ergebnis mit einem unterschiedlichen Validitätsbereich der beiden Tests, letztendlich also der Zielsetzung der Studien, die sich in der Entwicklung von Testaufgaben widerspiegelt.

Auch für den Bereich der Lesekompetenz liegen erste Analysen zum Zusammenhang zwischen nationalen und internationalen Kompetenztests vor. Für den Vergleich zwischen der Lesekompetenzmessung auf Basis der Bildungsstandards in der Grundschule und im Rahmen von IGLU berichten Pietsch, Böhme, Robitzsch und Stubbe (2009) inhaltliche Übereinstimmungen zwischen den verwendeten Kompetenzstufenmodellen. Die empirischen Zusammenhänge zwischen den Testverfahren liegen auf latenter Ebene bei $r = .84$. Die Stärke dieser latenten Zusammenhänge ist vergleichbar mit jener zwischen den PISA-Tests zu den drei Kompetenzbereichen (vgl. etwa die oben genannten Ergebnisse aus PISA 2003). Die Autorinnen und Autoren kommen zu dem Schluss, „dass ein unmittelbarer, linearer Vergleich der Ergebnisse aus IGLU 2006 und den länderübergreifenden Bildungsstandards nicht sinnvoll ist“ (Pietsch et al., 2009, S. 426). Sie vermuten, dass die geringe Korrelation auf deutliche Unterschiede in der Aufgabenentwicklung und in den verwendeten Aufgabentypen zwischen den beiden Studien, insbesondere aber auf die Verwendung verschiedener Arten von Stimulustexten zurückzuführen ist.

Anhand der hier dargestellten Ergebnisse wird deutlich, dass Tests aus unterschiedlichen Studien, auch wenn sie auf das gleiche Kompetenzkonstrukt abzielen, in der Regel nicht vollständig das Gleiche messen. Insbesondere bei der gemeinsamen Modellierung von Daten aus internationalen Studien mit denen aus nationalen Erhebungen erscheint der sprachliche beziehungsweise nationale Ursprung der Testaufgaben selbst als eine der Ursachen für die jeweils spezifischen Varianzanteile (Artelt & Baumert, 2004; Jurecka, 2010).

Aufbauend auf den hier dargestellten Befunden beschäftigt sich auch der vorliegende Artikel mit der Frage des empirischen Zusammenhangs zwischen Testwerten aus einem nationalen und einem internationalen Kompetenztest. Im Folgenden werden die Fragestellung, die Datengrundlage und das methodische Vorgehen dargestellt.

4. Methode

4.1 Fragestellungen

Der vorliegende Artikel geht der Frage nach, inwieweit die Kompetenzverteilungen der Schülerinnen und Schüler in Deutschland, die auf der Grundlage des Lesetests in PISA 2009 und des Lesetests im Rahmen des bildungsstandardbasierten Ländervergleichs 2009 ermittelt wurden, empirisch vergleichbar sind.

Diese Fragestellung lässt sich wie folgt weiter differenzieren:

- a) Wie unterscheiden sich die Verteilungen von Schülerinnen und Schülern, die beide Tests bearbeitet haben, auf den jeweiligen Kompetenzstufen der Lesekompetenz?
- b) Welcher korrelative Zusammenhang besteht zwischen den mit beiden Testverfahren gemessenen Kompetenzwerten?

4.2 Datengrundlage

Die Datengrundlage für die folgenden Analysen bildet eine Stichprobe von 8061 Schülerinnen (49.3%) und Schülern der 9. Jahrgangsstufe aus 201 Schulen in Deutschland. Dabei wurden alle Schularten des allgemeinbildenden Schulsystems einbezogen. Aufgrund des Erhebungsdesigns wurden Schülerinnen und Schüler aus beruflichen Schulen sowie aus Förderschulen in dieser Stichprobe allerdings nicht einbezogen. Das Alter der Schülerinnen und Schüler betrug im Durchschnitt 15 Jahre (Minimum 13, Maximum 19). Diese Stichprobe verband den bildungsstandardbasierten Ländervergleich Sprachen 2009 mit der Erhebung von PISA 2009. Für beide Studien wurden jeweils eigenständige Stichproben gezogen, wobei die hier verwendete gemeinsame Stichprobe lediglich einen Teil der Gesamtstichproben zur bildungsstandardbasierten Erhebung und von PISA darstellt (vgl. Abb. 1). Die Schülerinnen und Schüler bearbeiteten am ersten Testtag die internationalen PISA-Aufgaben und an einem zweiten Testtag die Testaufgaben des bildungsstandardbasierten Ländervergleichs.

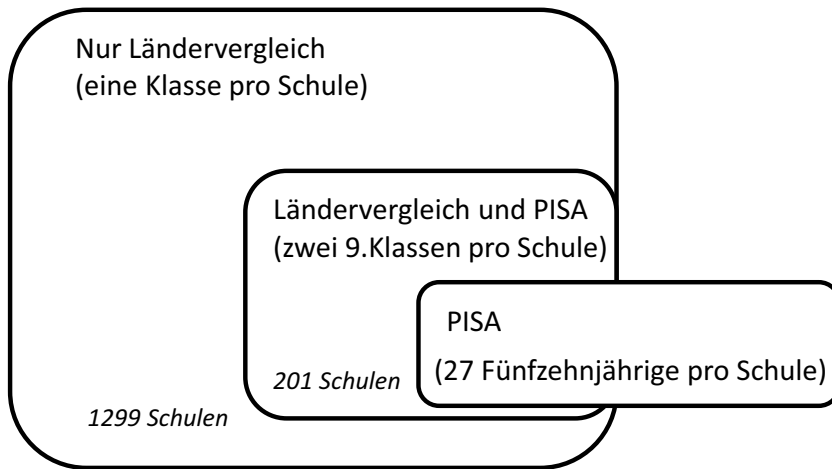


Abb. 1: Stichproben im Ländervergleich und der neunten Klassen in PISA 2009

4.3 Analytisches Vorgehen

Der Vergleich der empirischen Verteilungen der Schülerinnen und Schüler auf die Kompetenzstufen von PISA und den Bildungsstandards basiert auf den für beide Studien separat skalierten Testwerten. Pro Studie wurden für jede Person fünf *Plausible Values* (PVs) geschätzt (Köller et al., 2010; OECD, 2012). Anhand dieser PVs wurden für jede Studie separat die prozentualen Verteilungen der Stichprobe auf den Kompetenzstufen erzeugt und einander gegenübergestellt. Verwendet wurde hierfür die Software SPSS mit dem Zusatzmodul Macro für *Replicate Weights* (Balanced Repeated Replication, vgl. OECD, 2012, 2009d).

In diese Analyse gingen die Rohwerte von 101 Einzelaufgaben aus PISA sowie von 99 Einzelaufgaben zu den Bildungsstandards aus dem Kompetenzbereich Lesen ein. Zur Beurteilung der gemeinsamen dimensional Struktur der beiden Kompetenztests wurde die Passung von zwei Raschmodellen verglichen. Geschätzt wurden ein eindimensionales Raschmodell, bei dem alle Aufgaben auf einem gemeinsamen latenten Faktor luden, sowie ein zweidimensionales Raschmodell. Im zweidimensionalen Modell definierte die Zugehörigkeit der Items zu den Tests die Ladungsstruktur, d.h. alle Aufgaben des PISA-Lesetests luden ausschließlich auf einem Faktor und alle Aufgaben des Bildungsstandards-Lesetests luden ausschließlich auf einem zweiten Faktor (Adams, Wilson & Wang, 1997). Die latente Korrelation zwischen den beiden Faktoren wurde frei geschätzt.

In den Analysen wurden Schülergewichte zur Berücksichtigung unterschiedlicher Ziehungswahrscheinlichkeiten verwendet. Der Schachtelung der Schülerinnen und

Schüler in Schulen wurde durch die Verwendung der in Mplus implementierten Pseudo-Maximum-Likelihood-Schätzung (PML) für komplexe Stichproben (Asparouhov & Muthén, 2005) Rechnung getragen. Hierzu wurde die Software Mplus 6.1 verwendet (Muthén & Muthén, 2009). Vergleichskriterien für die Beurteilung der Modellpassung waren der AIC und BIC (Akaike, 1974; Schwarz, 1978).

5. Ergebnisse

5.1 Verteilung auf die Kompetenzstufen

Um zu vergleichen, wie sich die Personen der gemeinsamen Stichprobe auf die Kompetenzstufen von PISA und den Bildungsstandards verteilen, wurden die Verteilungen jeweils getrennt für die Kompetenzstufen nach PISA und die Kompetenzstufen der Bildungsstandards gebildet. Jede Person wurde also zweimal zugeordnet: Einmal anhand ihres Testwertes in PISA auf die Kompetenzstufen von PISA 2009 und einmal anhand des Testwertes, den sie in den Aufgaben auf Basis der Bildungsstandards erzielt hat, auf die Kompetenzstufen des Bildungsstandardmodells. Anhand dieser Zuordnung kann keine Aussage darüber getroffen werden, ob sich eine Person jeweils auf den gleichen Stufen in beiden Modellen befindet. Die entsprechenden Verteilungen auf den Kompetenzstufen sind vergleichend in Tabelle 4 gegenübergestellt.¹

Wie anhand der prozentualen Häufigkeiten ersichtlich wird, unterscheiden sich die Verteilungen der Personen auf die Kompetenzstufen nicht bedeutsam zwischen PISA und den Bildungsstandards. Bedingt durch die Unterschiede im Differenzierungsgrad der Kompetenzmodelle ist die Stufe VI als oberste Kompetenzstufe nur im Modell von PISA belegt (0.3 % der Stichprobe). Dem unteren Kompetenzbereich (Stufe I) werden im Modell der Bildungsstandards 16.3% der Stichprobe zugeordnet. Dieser Wert entspricht in etwa dem Anteil von Personen, die im Modell von PISA zusammen den Stufen Ib und Ia sowie dem Bereich unter Kompetenzstufe Ib zugeordnet werden (16.1%). PISA differenziert in diesem Bereich zusätzlich noch zwischen Personen, die Kompetenzstufe Ib nicht erreichen (0.5%), und jenen auf Kompetenzstufe Ib (3.4%).

1 Es ist zu beachten, dass die hier dargestellte Verteilung der Neuntklässlerinnen und Neuntklässler in Deutschland auf die Kompetenzstufen des Bildungsstandards-Modells weder mit den für die Normierungsstudie zu den Bildungsstandards berichteten Verteilungen vergleichbar ist, noch mit den in Köller et al. (2010) berichteten Verteilungen zum Ländervergleich 2009. Die dort wiedergegebenen Verteilungen beziehen sich nur auf jene Neuntklässlerinnen und Neuntklässler in Deutschland, die den Mittleren Schulabschluss anstreben, während die im vorliegenden Artikel berichtete Verteilung auch jene Neuntklässlerinnen und Neuntklässler einschließt, die den Hauptschulabschluss anstreben.

Kompetenzstufen	PISA		Bildungsstandards für den MSA	
	%	SE	%	SE
VI	0.3	0.1	5.4	0.5
V	5.2	0.4		
IV	21.7	1.0	19.3	0.9
III	31.4	0.8	31.9	0.8
II	25.3	1.0	27.0	1.0
Ia	12.2	0.7	16.4	1.0
Ib	3.4	0.4		
unter Ib	0.5	0.1		

Tab. 4: Verteilung der Neuntklässlerinnen und Neuntklässler auf die Kompetenzstufen von PISA 2009 und des Kompetenzstufenmodells zu den Bildungsstandards für den MSA im Fach Deutsch, Bereich Lesen (prozentuale Häufigkeiten und Standardfehler [SE])

5.2 Zusammenhang zwischen den Lesetests

Um zu prüfen, welcher empirische Zusammenhang zwischen dem Lesetest von PISA und dem Lesetest zu den Bildungsstandards besteht, wurden auf Basis der Ergebnisse

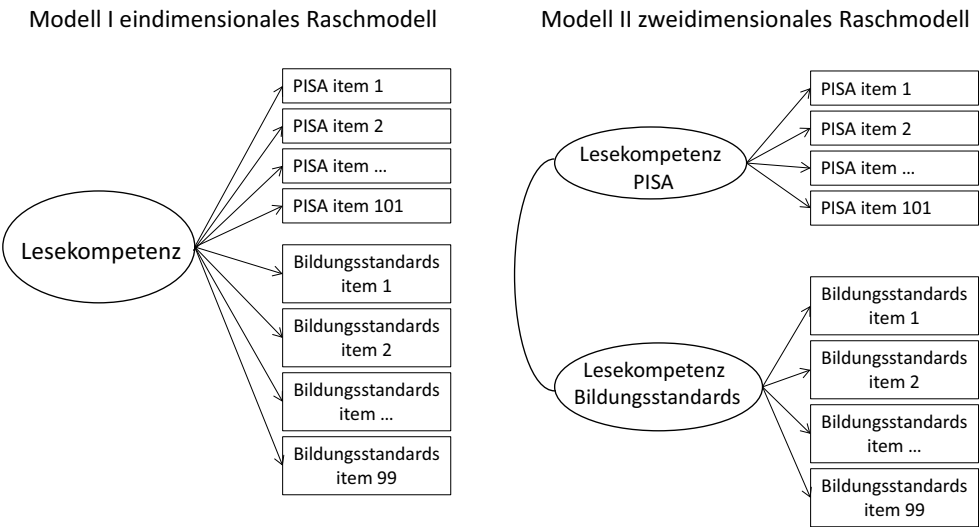


Abb. 2: Eindimensionale und zweidimensionale Raschmodellierung des PISA-Lesetests und des auf den Bildungsstandards basierenden Lesetests (schematische Darstellung)

der Schülerinnen und Schüler in den einzelnen Testaufgaben zwei Modelle spezifiziert und verglichen (vgl. Abb. 2).

Das erste Modell geht von einem gemeinsamen Faktor der Lesekompetenz aus, auf dem alle Testaufgaben aus PISA und dem Ländervergleich gleichermaßen laden. Das zweite Modell spezifiziert zwei korrelierte latente Faktoren der Lesekompetenz – einen für PISA und einen für die Bildungsstandards-Aufgaben –, auf denen jeweils nur die Aufgaben der zugehörigen Studie laden. Beim Vergleich der beiden Modelle auf Basis des AIC und BIC erweist sich die Passung des zweidimensionalen Modells (AIC = 473154; BIC = 474623) für die Beschreibung der in den Daten vorliegenden Zusammenhänge der Passung des eindimensionalen Modells (AIC = 474706; BIC = 476161) als überlegen. Die latente Korrelation von $r = .85$ zwischen den Faktoren im zweidimensionalen Modell kann dahingehend interpretiert werden, dass es sich um zwei korrelierte Konstrukte handelt, die jeweils einen eigenen Beitrag zur Erklärung der in den Schülerantworten auftretenden Varianz liefern.

6. Diskussion

Der vorliegende Artikel untersuchte die Frage, inwieweit die auf den Testergebnissen basierenden Kompetenzverteilungen der Schülerinnen und Schüler in Deutschland zwischen dem PISA-Lesetest und dem Lesetest auf Basis der Bildungsstandards für den Mittleren Schulabschluss empirisch vergleichbar sind. Dazu wurden die theoretischen Definitionen der Lesekompetenz von PISA und der Bildungsstandards, der Prozess der Aufgabenentwicklung sowie die Kompetenzmodelle beider Studien vergleichend dargestellt. Darüber hinaus wurden auf Basis einer gemeinsamen Stichprobe die empirischen Zusammenhänge zwischen den Ergebnissen der Leistungstests bestimmt und die Häufigkeitsverteilungen auf den Kompetenzstufen verglichen.

Die Analyse des empirischen Zusammenhangs zwischen den Kompetenzkonstrukten in PISA und der Studie auf Basis der Bildungsstandards hat gezeigt, dass ein zweidimensionales Modell, das separate korrelierte Kompetenzdimensionen annimmt, die in den Daten vorhandene Struktur besser abbildet als ein eindimensionales Modell. Der empirische Zusammenhang zwischen den Testwerten aus der zweidimensionalen Skalierung beträgt $r = .85$. Die latente Korrelation zwischen beiden Tests liegt damit in einem ähnlichen Bereich wie der Zusammenhang zwischen dem Lesetest auf Basis der Bildungsstandards im Fach Deutsch für den Primarbereich und dem IGLU-Lesetest. Zudem ist diese Korrelation vergleichbar mit Zusammenhängen zwischen unterschiedlichen Kompetenzbereichen in PISA, also beispielsweise zwischen Lesen und Mathematik. Die latente Korrelation ist sogar geringer als der empirische Zusammenhang zwischen dem nationalen Test zu den Bildungsstandards-Mathematik und dem internationalen PISA-Mathematiktest.

Dieses Ergebnis wird im Folgenden vor dem Hintergrund der zugrunde liegenden Kompetenzkonzepte, der Zielsetzungen der Studien, der jeweiligen Prozesse der Aufgabenentwicklung und der Gestaltung der Kompetenzmodelle diskutiert.

Kompetenzkonzepte und Zielsetzungen der Studien

Das Konstrukt der Lesekompetenz kann konzeptuell als mehrdimensional angesehen werden (Artelt & Schlagmüller, 2004). PISA berücksichtigt verschiedene Aspekte der Lesekompetenz dadurch, dass diese als fünf getrennte Teildimensionen gemessen und berichtet werden. Die Ergebnisse zeigen, dass die Mittelwertsunterschiede zwischen Staaten in den verschiedenen Aspekten der Lesekompetenz beträchtlich sind (Naumann et al., 2010). Dies deutet darauf hin, dass die unterschiedlichen Teildimensionen der Lesekompetenz in unterschiedlichem Maße zur Gesamtkompetenz beitragen können.

Auch im theoretischen Konzept der Bildungsstandards wird der Mehrdimensionalität des Lesekompetenzkonstrukts Rechnung getragen, indem Anforderungsniveaus und Textsorten differenziert werden (vgl. KMK, 2004a). Diese Aspekte werden zwar bei der Testkonstruktion berücksichtigt, es erfolgt jedoch keine nach Teildimensionen getrennte Skalierung oder Berichtlegung. Analysen zur Dimensionalität der Lesekompetenz in den Tests zu den Bildungsstandards stehen noch aus. Vor diesem Hintergrund kann der vorliegende Artikel beim Vergleich beider Studien nur den über alle Items hinweg skalierten Gesamtestwert berücksichtigen. Es ist anzunehmen, dass die geringe Korrelation zwischen den beiden Lesetests zumindest teilweise durch die Mehrdimensionalität der Kompetenzkonstrukte erklärt werden kann. In weiterführenden Analysen sollte daher die Dimensionalität des Lesetests der Bildungsstandards empirisch überprüft werden.

Testentwicklung und Testaufgaben

Die Testentwicklung auf Grundlage der Bildungsstandards erfolgte mit nationaler Ausrichtung und dem Ziel, Testaufgaben für die Verwendung im Ländervergleich innerhalb von Deutschland bereitzustellen. Die PISA-Aufgabenentwicklung zielt hingegen darauf ab, eine in allen internationalen Teilnehmerstaaten vergleichbare Messung von Lesekompetenz zu ermöglichen. Das verwendete Aufgabenmaterial ist entsprechend breit, entstammt unterschiedlichen Kulturkreisen und ist nicht explizit an Curricula oder Bildungsstandards orientiert.

Dementsprechend finden sich zwischen den Studien auch Unterschiede in den Testaufgaben. Beispielsweise erfordern die Aufgaben zu den Bildungsstandards Kenntnisse über literarische Textgattungen, die in Deutschland im Unterricht der Sekundarstufe I behandelt werden. Die Aufgaben des PISA-Tests beziehen sich hingegen stärker auf den Umgang mit Alltagstexten, wie sie weltweit in jeder Kultur vorkommen könnten. Ein wichtiger formaler Unterschied besteht auch in der Länge der Stimulustexte: Während die Texte in PISA in der Regel eine Seite nicht überschreiten, wurden im bildungsstandardbasierten Ländervergleich überwiegend Texte mit einer Länge von mehr als einer Seite eingesetzt. Zudem verwendet PISA einen höheren Anteil an Aufgaben mit offenem Antwortformat.

Bei der gemeinsamen Skalierung aller Testaufgaben, wie sie im vorliegenden Artikel vorgenommen wurde, spielen diese Aufgabeneigenschaften eine wesentliche Rolle. Da-

her könnte die bessere Passung des zweidimensionalen Modells auf die Daten durchaus auch auf die unterschiedliche Genese der jeweils verwendeten Testaufgaben zurückzuführen sein.

Verteilung auf den Kompetenzstufen

Die Verteilung auf den Kompetenzstufen unterscheidet sich zwischen PISA und den Bildungsstandards nicht wesentlich. Die prozentualen Anteile an Personen auf den jeweiligen Stufen sind vergleichbar. Erklärt werden kann dieses Ergebnis durch das methodische Vorgehen bei der Bestimmung der Cut-off-Werte für die Kompetenzstufen im Modell der Bildungsstandards. Das Standard-Setting orientierte sich hier explizit an den Stufenbeschreibungen des PISA-Modells und der Stichprobenverteilung der Neuntklässler in Deutschland, wie sie in PISA 2000 ermittelt wurde. Daher war zu erwarten, dass sich die Verteilungen auf den Kompetenzstufen zu den Bildungsstandards, die anhand von Daten der Normierungsstudie festgelegt wurden, nicht wesentlich von der Verteilung für PISA 2009 unterscheiden.

Unterschiede in der Verteilung entstehen lediglich dadurch, dass das Kompetenzstufenmodell von PISA 2009 stärker differenziert als das Modell von PISA 2000 und den Bildungsstandards. Werden die zusätzlich differenzierten Stufen bei PISA allerdings zusammengefasst (die Stufen Ib und Ia sowie die Stufen V und VI) und den jeweils korrespondierenden Stufen der Bildungsstandards gegenübergestellt, zeigen sich für keine der Stufen signifikante Unterschiede in den prozentualen Häufigkeiten zwischen den beiden Studien. Wie die relativ geringe latente Korrelation im zweidimensionalen Modell zeigt, bedeutet dies jedoch nicht zwangsläufig, dass sich dieselben Personen in beiden Studien auf vergleichbaren Kompetenzstufen befinden (vgl. auch Pietsch et al., 2009).

Fazit und Ausblick

Insgesamt lässt sich auf Basis der hier dargestellten Ergebnisse für den Vergleich der Testergebnisse von PISA und zum bildungsstandardbasierten Test feststellen, dass a) die empirischen Verteilungen der Schülerinnen und Schüler in Deutschland auf den jeweiligen Kompetenzstufen prozentual vergleichbar sind, jedoch b) die in beiden Studien erfassten Lesekompetenzen empirisch als korrelierte, gleichzeitig jedoch auch distinkte Konstrukte zu betrachten sind.

Anhand eines erzielten Kompetenzwerts oder einer erreichten Kompetenzstufe in PISA kann also nicht direkt darauf geschlossen werden, welchen Kompetenzwert oder welche Kompetenzstufe dieselbe Person in der Studie auf Basis der Bildungsstandards erreicht hätte, und umgekehrt.

Die wesentlichen Unterschiede in den Zielsetzungen der Studien, in den Konstruktdefinitionen sowie in der Ausgestaltung des Testmaterials bieten mögliche Erklärungen für die vorliegenden Ergebnisse. Diese Ergebnisse reihen sich ein in die bisheri-

gen Ergebnisse zum Vergleich national entwickelter Testaufgaben mit Kompetenztests aus internationalen Large Scale Assessments.

Ausgehend von diesen Befunden und Überlegungen lassen sich mehrere Fragestellungen für weiterführende Analysen identifizieren. Zum einen sollte – wie bereits erwähnt – untersucht werden, ob die Aufgaben auf Basis der Bildungsstandards eine mehrdimensionale Struktur aufweisen, mit der sich wie bei PISA Teildimensionen der Lesekompetenz aufzeigen lassen. Anhand dieser Informationen ließen sich auch komplexere Modelle der Lesekompetenz mit den Daten beider Studien analysieren. Zur weiteren Erklärung der Konstruktdimensionalität empfiehlt es sich, eine Einschätzung der verwendeten Testaufgaben durch einschlägige Experten vornehmen zu lassen. Wie bereits dargestellt wurde, ist davon auszugehen, dass die Testaufgaben aufgrund ihrer unterschiedlichen Entstehungskontexte zu einem Teil auf unterschiedliche Aspekte der Lesekompetenz abzielen. Darüber hinaus könnten anhand von Experteneinschätzungen die jeweils korrespondierenden Kompetenzstufen beider Kompetenzstufenmodelle inhaltlich verglichen werden. Dies könnte Aufschluss darüber geben, warum zwar die Häufigkeitsverteilung auf den Kompetenzstufen weitgehend vergleichbar ist, die erfassten Kompetenzen jedoch nicht identisch sind.

Literatur

- Adams, R., Wilson, M., & Wang, W. (1997). The Multidimensional Random Coefficients Multinomial Logit Model. *Applied Psychological Measurement*, 21(1), 1-23.
- Adams, R., & Wu, M. (Hrsg.) (2002). *PISA 2000 Technical Report*. Paris: OECD.
- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6), 716-723.
- Artelt, C., & Baumert, J. (2004). Zur Vergleichbarkeit von Schülerleistungen bei Leseaufgaben unterschiedlichen sprachlichen Ursprungs. *Zeitschrift für Pädagogische Psychologie*, 18(4), 171-185.
- Artelt, C., Naumann, J., & Schneider, W. (2010). Lesemotivation und Lernstrategien. In E. Klieme, C. Artelt, J. Hartig, N. Jude, O. Köller, M. Prenzel, W. Schneider & P. Stanat (Hrsg.), *PISA 2009. Bilanz nach einem Jahrzehnt* (S. 73-112). Münster: Waxmann.
- Artelt, C., & Schlagmüller, M. (2004). Der Umgang mit literarischen Texten als Teilkompetenz im Lesen – Dimensionsanalysen und Ländervergleiche. In U. Schiefele, C. Artelt, W. Schneider & P. Stanat (Hrsg.), *Struktur, Entwicklung und Förderung von Lesekompetenz – Vertiefende Analysen im Rahmen von PISA 2000* (S. 169-196). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Artelt, C., Stanat, P., Schneider, W., & Schiefele, U. (2001). Lesekompetenz. Testkonzeption und Ergebnisse. In Deutsches PISA-Konsortium (Hrsg.), *PISA 2000* (S. 69-131). Opladen: Leske + Budrich.
- Asparouhov, T., & Muthén, B. O. (2005). *Multivariate Statistical Modeling with Survey Data* (Proceedings of the Federal Committee on Statistical Methodology (FCSM) Research Conference). http://www.fcsm.gov/05papers/Asparouhov_Muthen_IIA.pdf [26.11.2012].
- Autorengruppe Bildungsberichterstattung (2012). *Bildung in Deutschland 2012. Ein indikatoren-gestützter Bericht mit einer Analyse zur kulturellen Bildung im Lebenslauf*. Bielefeld: wbv.
- Baumert, J., Artelt, C., Klieme, E., Neubrand, M., Prenzel, M., Schiefele, U., Schneider, W., Tillmann, K. J., & Weiß, M. (Hrsg.) (2002). *PISA 2000. Die Länder der Bundesrepublik Deutschland im Vergleich*. Opladen: Leske + Budrich.

- Baumert, J., Bos, W., & Lehmann, R. (Hrsg.) (2000). *TIMSS/III: Dritte Internationale Mathematik- und Naturwissenschaftsstudie – Mathematische und naturwissenschaftliche Bildung am Ende der Schullaufbahn*. Opladen: Leske + Budrich.
- Blum, W., Neubrand, M., Ehmke, T., Senkbeil, M., Jordan, A., Ulfig, F., & Carstensen, C. H. (2004). Mathematische Kompetenz. In PISA-Konsortium Deutschland (Hrsg.), *PISA 2003. Der Bildungsstand der Jugendlichen in Deutschland – Ergebnisse des zweiten internationalen Vergleichs* (S. 47-92). Münster: Waxmann.
- Böhme, K., Leucht, M., Schipolowski, S., Porsch, R., Knigge, M., & Köller, O. (2010). Anlage und Durchführung des Ländervergleichs. In O. Köller, M. Knigge & B. Tesch (Hrsg.), *Sprachliche Kompetenzen im Ländervergleich* (S. 65-85). Münster: Waxmann.
- Böhme, K., Neumann, D., & Schipolowski, S. (2010). Beschreibung der im Ländervergleich im Fach Deutsch untersuchten Kompetenzen. In O. Köller, M. Knigge & B. Tesch (Hrsg.), *Sprachliche Kompetenzen im Ländervergleich* (S. 19-24). Münster: Waxmann.
- Bos, W., Hornberg, S., Arnold, K.-H., Faust, G., Fried, L., Lankes, E.-M., Schwippert, K., & Valtin, R. (Hrsg.) (2007). *IGLU 2006. Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich*. Münster: Waxmann.
- Bremerich-Vos, A., Behrens, U., Böhme, K., Krelle, M., Neumann, D., Robitzsch, A., Schipolowski, S., & Köller, O. (2010). Kompetenzstufenmodelle für das Fach Deutsch. In O. Köller, M. Knigge & B. Tesch (Hrsg.), *Sprachliche Kompetenzen im Ländervergleich* (S. 37-49). Münster: Waxmann.
- Granzer, D. (2009). Von Bildungsstandards zu ihrer Überprüfung: Grundlagen der Item- und Testentwicklung. In D. Granzer, O. Köller & A. Bremerich-Vos (Hrsg.), *Bildungsstandards Deutsch und Mathematik. Leistungsmessung in der Grundschule* (S. 21-30). Weinheim/Basel: Beltz.
- Groeben, N., & Hurrelmann, B. (Hrsg.) (2004). *Lesesozialisation in der Mediengesellschaft: Ein Forschungsüberblick*. Weinheim: Juventa.
- Guthrie, J. T., & Wigfield, A. (2000). Engagement and Motivation in Reading. In M. L. Kamil, P. Mosenthal, P. D. Pearson & R. Barr (Hrsg.), *Handbook of Reading Research*. Mahwah: Lawrence Erlbaum Associates.
- Hartig, J., & Frey, A. (2012). Validität des Tests zur Überprüfung des Erreichens der Bildungsstandards in Mathematik. Zusammenhänge mit den bei PISA gemessenen Kompetenzen und Varianz zwischen Schulen und Schulformen. *Diagnostica*, 58, 3-14.
- Hartig, J., & Jude, N. (2007). Empirische Erfassung von Kompetenzen und psychometrische Kompetenzmodelle. In J. Hartig & E. Klieme (Hrsg.), *Möglichkeiten und Voraussetzungen technologiebasierter Kompetenzdiagnostik* (S. 17-36). Bonn/Berlin: Bundesministerium für Bildung und Forschung.
- Hartig, J., & Klieme, E. (2006). Kompetenz und Kompetenzdiagnostik. In K. Schweizer (Hrsg.), *Leistung und Leistungsdiagnostik* (S. 127-143). Berlin: Springer.
- Jude, N., & Klieme, E. (2010). Das Programme for International Student Assessment (PISA). In E. Klieme, C. Artelt, J. Hartig, N. Jude, O. Köller, M. Prenzel, W. Schneider & P. Stanat (Hrsg.), *PISA 2009. Bilanz nach einem Jahrzehnt* (S. 11-20). Münster: Waxmann.
- Jurecka, A. (2010). *Zum Zusammenhang von Differentiellen Item Funktionen und Testkultur*. Frankfurt a.M.: Universitätsbibliothek Johann Christian Senckenberg.
- Klieme, E., Artelt, C., Hartig, J., Jude, N., Köller, O., Prenzel, M., Schneider, W., & Stanat, P. (Hrsg.) (2010). *PISA 2009. Bilanz nach einem Jahrzehnt*. Münster: Waxmann.
- Klieme, E., Eichler, W., Helmke, A., Lehmann, R. H., Nold, G., Rolff, H.-G., Schröder, K., Thomé, G., & Willenberg, H. (2008). *Unterricht und Kompetenzerwerb in Deutsch und Englisch. DESI-Ergebnisse Band 2*. Weinheim/Basel: Beltz.
- Klieme, E., Neubrand, M., & Lüdtke, O. (2001). Mathematische Grundbildung: Testkonzeption und Ergebnisse. In Deutsches PISA-Konsortium (Hrsg.), *PISA 2000. Basiskompetenzen*

- von Schülerinnen und Schülern im internationalen Vergleich (S. 139-190). Opladen: Leske + Budrich.
- KMK (1997). *Grundsätzliche Überlegungen zu Leistungsvergleichen innerhalb der Bundesrepublik Deutschland. Konstanzer Beschluss*. Bonn/Berlin: KMK.
- KMK (2004a). *Bildungsstandards im Fach Deutsch für den Mittleren Schulabschluss (Jahrgangsstufe 10)*. München: Luchterhand.
- KMK (2004b). *Bildungsstandards der Kultusministerkonferenz: Erläuterungen zur Konzeption und Entwicklung*. München: Luchterhand.
- KMK (2006). *Gesamtstrategie der Kultusministerkonferenz zum Bildungsmonitoring*. München: Wolters Kluwer.
- Köller, O., Knigge, M., & Tesch, B. (2010). *Sprachliche Kompetenzen im Ländervergleich*. Münster: Waxmann.
- Leutner, D., Klieme, E., Meyer, K., & Wirth, J. (2004). Problemlösen. In PISA-Konsortium Deutschland (Hrsg.), *PISA 2003: Der Bildungsstand der Jugendlichen in Deutschland – Ergebnisse des zweiten internationalen Vergleichs* (S. 147-175). Münster: Waxmann.
- Mildner, D., Hochweber, J., & Frey, A. (2013). Vergleichende Analysen der Kompetenzen von Fünfzehnjährigen und Neuntklässlern in den deutschen PISA-Erhebungen 2003 bis 2009. *Zeitschrift für Pädagogik*, 59. Beiheft, 151-157.
- Muthén, L. K., & Muthén, B. O. (2009). *Mplus User's Guide* (6. Aufl.). Los Angeles: Muthén & Muthén.
- Naumann, J., Artelt, C., Schneider, W., & Stanat, P. (2010). Lesekompetenz von PISA 2000 bis PISA 2009. In E. Klieme, C. Artelt, J. Hartig, N. Jude, O. Köller, M. Prenzel, P. Stanat & W. Schneider (Hrsg.), *PISA 2009. Bilanz nach einem Jahrzehnt* (S. 23-71). Münster: Waxmann.
- OECD (2009a). *PISA 2009 Assessment Framework – Key Competencies in Reading, Mathematics and Science*. Paris: OECD.
- OECD (2009b). *PISA Take the Test: Sample Questions from OECD's PISA Assessments*. Paris: OECD.
- OECD (2009c). *PISA 2006 Technical Report*. Paris: OECD.
- OECD (2009d). *PISA Data Analysis Manual: SPSS® Second edition*. Paris: OECD.
- OECD (2010). *PISA 2009 Ergebnisse: Was Schülerinnen und Schüler wissen und können. Schülerleistungen in Lesekompetenz, Mathematik und Naturwissenschaften* (Band I). Berlin: wbv.
- OECD (2012). *PISA 2009 Technical Report*. Paris: OECD.
- Pant, H. A., Tiffin-Richards, S. P., & Köller, O. (2010). Standard-Setting für Kompetenztests im Large-Scale-Assessment. *Zeitschrift für Pädagogik*, 56. Beiheft, 175-188.
- Pietsch, M., Böhme, K., Robitzsch, A., & Stubbe, T. (2009). Das Stufenmodell zur Lesekompetenz der länderübergreifenden Bildungsstandards im Vergleich zu IGLU 2006. In A. Bremerich-Vos, D. Granzer & O. Köller (Hrsg.), *Bildungsstandards Deutsch und Mathematik. Leistungsmessung in der Grundschule* (S. 393-416). Weinheim/Basel: Beltz.
- Pöhlmann, C., Neumann, D., Tesch, B., & Köller, O. (2010). Die Bildungsstandards im allgemeinbildenden Schulsystem. In O. Köller, M. Knigge & B. Tesch (Hrsg.), *Sprachliche Kompetenzen im Ländervergleich* (S. 13-18). Münster: Waxmann.
- Prenzel, M., Artelt, C., Baumert, J., Blum, W., Hammann, M., Klieme, E., & Pekrun, R. (Hrsg.) (2008). *PISA 2006 in Deutschland. Die Kompetenzen der Jugendlichen im dritten Ländervergleich*. Münster: Waxmann.
- Prenzel, M., Baumert, J., Blum, W., Lehmann, R. H., Leutner, D., Neubrand, M., Pekrun, R., Rost, J., & Schiefele, U. (Hrsg.) (2005). *PISA 2003. Der zweite Vergleich der Länder in Deutschland – Was wissen und können Jugendliche?* Münster: Waxmann.
- Prenzel, M., Rost, J., Senkbeil, M., Häußler, P., & Klopp, A. (2001). Naturwissenschaftliche Grundbildung. In Deutsches PISA-Konsortium (Hrsg.), *PISA 2000. Basiskompetenzen von*

Schülerinnen und Schülern im internationalen Vergleich (S. 191-248). Opladen: Leske + Budrich.

Schwarz, G. E. (1978). Estimating the dimension of a model. *Annals of Statistics*, 6(2), 461-464.
Stanat, P., Pant, H. A., Böhme, K., & Richter, D. (Hrsg.) (2012). *Ergebnisse des IQB-Ländervergleichs 2011. Kompetenzen von Schülerinnen und Schülern am Ende der vierten Jahrgangsstufe in den Fächern Deutsch und Mathematik*. Münster: Waxmann.

Anschrift der Autor(inn)en

Dr. Nina Jude, Deutsches Institut für Internationale Pädagogische Forschung (DIPF),
Schloßstr. 29, 60486 Frankfurt am Main, Deutschland
E-Mail: jude@dipf.de

Prof. Dr. Johannes Hartig, Deutsches Institut für Internationale Pädagogische Forschung
(DIPF), Schloßstr. 29, 60486 Frankfurt am Main, Deutschland
E-Mail: hartig@dipf.de

Dipl.-Psych. Stefan Schipolowski, Humboldt-Universität zu Berlin, Institut zur Qualitäts-
entwicklung im Bildungswesen (IQB), Unter den Linden 6, 10099 Berlin, Deutschland
E-Mail: Stefan.Schipolowski@IQB.hu-berlin.de

Dr. Katrin Böhme, Humboldt-Universität zu Berlin, Institut zur Qualitätsentwicklung im
Bildungswesen (IQB), Unter den Linden 6, 10099 Berlin, Deutschland
E-Mail: Katrin.Boehme@IQB.hu-berlin.de

Prof. Dr. Petra Stanat, Humboldt-Universität zu Berlin, Institut zur Qualitätsentwicklung im
Bildungswesen (IQB), Unter den Linden 6, 10099 Berlin, Deutschland
E-Mail: iqboffice@IQB.hu-berlin.de

Eckhard Klieme/Svenja Vieluf

Schulische Bildung im internationalen Vergleich

Ein Rahmenmodell für Kontextanalysen in PISA

1. Eine Meta-Perspektive auf PISA

Das *Programme for International Student Assessment* (PISA) baut auf der mehr als fünfzigjährigen Tradition der international vergleichenden empirischen Bildungsforschung auf. 1958 schlossen sich Wissenschaftlerinnen und Wissenschaftler aus Europa und den USA in der *International Association for the Evaluation of Educational Achievement* (IEA) zusammen, um die qualitativ und historisch-systematisch argumentierende Komparatistik durch Schulleistungsstudien zu ergänzen. Seit den 1960er Jahren führt die IEA regelmäßig internationale Leistungsstudien durch, derzeit u.a. die Primarschulstudien TIMSS und PIRLS (in Deutschland IGLU genannt). Über die Ziele und Grenzen dieses Forschungsparadigmas wird in der IEA seit langem und immer wieder neu diskutiert. So unterscheiden Keeves und Lietz (2011) zwei Modi der international vergleichenden Schulleistungsforschung: einen deskriptiven, auf die fortlaufende Beobachtung von Systemmerkmalen und Ergebnissen bezogenen Modus, und einen erklärenden Modus, der die Frage in den Vordergrund stellt, inwieweit bildungs- und erziehungswissenschaftliche Befunde auf unterschiedliche kulturelle und systemische Kontexte übertragbar sind. Dem internationalen Vergleich kommt in den beiden Modi eine höchst unterschiedliche Funktion zu: Das sogenannte Bildungsmonitoring, einschließlich „Ranking“ und „Benchmarking“, einerseits, die Ausweitung des Erkenntnishorizonts durch eine interkulturelle und systemische Dimension andererseits.

Die Chance, aber auch die Problematik der internationalen Leistungsstudien besteht darin, dass die Untersuchungen beide Arten des Zugriffs zwanglos ermöglichen. Dementsprechend werden zwei Arten von „Produkten“ für Politik, Öffentlichkeit, Profession und Forschung bereitgestellt: (1) *Indikatoren*, mithilfe derer sich Struktur und Funktion sowie Produktivität und Chancengleichheit in Bildungssystemen beobachten lassen. (2) Zuverlässige *Daten*, die es Vertretern unterschiedlicher Disziplinen ermöglichen, Grundlagenforschung zu betreiben, aber auch bildungspolitisch relevante Fragen zu untersuchen, also *Wissen* zu schaffen über Faktoren, die die Qualität von Bildungssystemen beeinflussen.

In der Öffentlichkeit wurde und wird fast ausschließlich der erste Modus rezipiert. So führten bereits in den 1970er Jahren die im Rahmen des Six Subject Survey (Walker, 1976) ermittelten kritischen Befunde zum Leistungsstand im deutschen Bildungssystem zu Konflikten zwischen Kultusministerkonferenz und Wissenschaft und letztlich zum Ausscheiden Deutschlands aus den internationalen Vergleichen. Die TIMSS-Mittelstufenstudie, an der Mitte der 1990er Jahre auch Deutschland wieder teilnahm, bestätigte bezüglich des Leistungsvergleichs im Kern den zwanzig Jahre zuvor ermittelten

Befund (Baumert et al., 1997), gab aber – vor allem durch die angeschlossene Videostudie – zusätzliche Impulse im Sinne des zweiten Interpretationsmodus, d.h. Erkenntnisse und weiterführende Anregungen für eine fachdidaktisch, unterrichtswissenschaftlich und kulturvergleichend ausgerichtete Forschung, die über das Modellprogramm SINUS wiederum den mathematischen und naturwissenschaftlichen Unterricht in Deutschland nachhaltig beeinflusst hat. Eine ähnliche Rolle spielt in jüngster Zeit die IEA-Studie TEDS-M zu Leistungen und Einstellungen von Lehramtsstudierenden (Blömeke, Kaiser & Lehmann, 2010).

Ende der 1990er Jahre begann die *Organisation für wirtschaftliche Zusammenarbeit und Entwicklung* (OECD) diese Forschungstradition zu nutzen, um ihr System statistischer Indikatoren zu erweitern. PISA erhebt seitdem alle drei Jahre in den Mitgliedsstaaten der OECD und in mittlerweile mehr als 30 Partnerstaaten Kompetenzen von 15-jährigen Schülerinnen und Schülern in den Bereichen Lesen, Mathematik und Naturwissenschaften sowie zusätzliche fächerübergreifende Kompetenzen wie etwa das Problemlösen (PISA 2003 mit schriftlichen Aufgaben, PISA 2012 mit computer-simulierten Szenarien und PISA 2015 mit Aufgaben zur Interaktion mit einem simulierten Partner). PISA liefert zuverlässige Informationen zur Beobachtung und für den Vergleich von Bildungssystemen. Mithilfe der Ergebnisse können Strukturen von Bildungssystemen beschrieben sowie deren Funktionieren und (soweit international messbar) deren Ergebnisse beobachtet werden. Beispielsweise beschreibt PISA Schülerleistungen, aber auch nicht-kognitive *Outcomes* (z.B. Motivation), Bildungswege von Schülerinnen und Schülern, Merkmale von Schulen wie Schulklima und Lehrerkoooperation sowie Aspekte der Chancengleichheit im Bildungswesen (z.B. die geschlechtsspezifische Bildungsbeteiligung und die Koppelung zwischen sozioökonomischem Status und Schülerleistungen). Damit stellt die Studie „Transparenz über Rahmenbedingungen, Prozesse und Ergebnisse des Bildungssystems her, nutzt dabei verschiedene Perspektiven und dient somit der öffentlichen Verantwortung für das Bildungsgeschehen“ (Klieme, 2011).

Anders als die Studien der IEA ist PISA sogar selbst eine Instanz, die zum Gegenstand gesellschafts- und erziehungswissenschaftlicher Forschung wurde. Eine internationale Autorengruppe beschreibt PISA als „wissensbasiertes Regulationsinstrument im bildungspolitischen Raum“ (Ozga, 2012, S. 169; Übersetzung durch die Autoren), dessen Wirkung darin bestehe, dass Daten in Wissen und schließlich in normative politische Argumente transformiert würden, die Problemlagen definieren oder sogar Empfehlungen für Problemlösungen geben. Für nationale Bildungspolitik bedeutet dies – je nach Sichtweise – neue Möglichkeiten zur Analyse der Bildungswirklichkeit und eine Ausweitung der politischen Steuerungsmittel (so Ozga, 2008), oder die Entmachtung durch eine „transnationale Koalition aus Forschern, Managern und Unternehmensberatern“ (Münch, 2009, S. 30). Tatsächlich werden die PISA-Studien gern als Legitimationsinstanz für das „borrowing“, d.h. die Übertragung bildungspolitischer Konzepte eines „erfolgreichen“ Landes auf andere Staaten, verwendet – seien es Marktmechanismen, Standards und Evaluationen, Lehrerbildung oder Ganztags- und Einheitsschul-

system. Diese Kritik trifft PISA härter als IEA-Studien, nicht zuletzt weil mit der OECD eine mächtige supranationale Institution der Auftraggeber ist.

Auch PISA ist aber mehr als ein Vergleichstest zur Identifizierung „erfolgreicher“ Bildungssysteme im Sinne des Leistungsrankings. Neben den kognitiven Kompetenzen werden sogenannte nicht-kognitive Outcomes erhoben, also erzieherische Wirkungen im Sinne von Fend (2006) wie etwa Lernmotivation, fachbezogene Einstellungen, kooperative vs. kompetitive Orientierungen und Verhaltensweisen im Bereich Schulschwänzen. Zunehmend wichtiger ist über die Erhebungen hinweg – auch auf internationaler Ebene – die Kontextualisierung der Testergebnisse geworden, die auf Befragungen von Schülerinnen und Schülern, Schulleitungen, Eltern und bei PISA 2015 erstmals auch Lehrkräften beruht. Hier werden Faktoren abgefragt, von denen aus der Schul- und Unterrichtsforschung bekannt ist, dass sie einen Effekt auf schulische Lernergebnisse, Einstellungen und Verhaltensweisen haben. Wichtige Ausgangsbedingungen der schulischen Arbeit sind beispielsweise die familiäre Lernumgebung sowie die Ausstattung und Finanzierung der Schule. Zu den erhobenen Prozessmerkmalen des Lehr- und Lerngeschehens gehören unter anderem individuelle Lernstrategien, die Unterrichtsqualität und das Schulklima. Dabei wird berücksichtigt, dass diese Merkmale auf verschiedenen Ebenen von Schulsystemen (Schüler-, Klassen-,¹ Schul- und Systemebene) anzusiedeln sind. PISA stellt somit eine hoch komplexe Datenbasis bereit, die – entsprechend der oben dargestellten Dualität der Schulleistungsstudien – sowohl für Zwecke des Bildungsmonitorings genutzt werden kann als auch für wissenschaftliche Untersuchungen mit allgemeinen, nationalen und international vergleichenden Fragestellungen.

Diese wissenschaftlichen Untersuchungen zielen, abstrakt gesprochen, auf Muster von Zusammenhängen zwischen Ausgangsbedingungen, Prozessmerkmalen und verschiedenen Lernergebnissen. Der erste Analyseschritt ist hierbei die Zerlegung der Varianz von Schülerleistungen und anderen Merkmalen auf die Schüler-, die Schul- und die Systemebene; sie liefert Hinweise, inwieweit bildungsbezogene Prozesse und Ergebnisse tatsächlich auf der Ebene der Einzelschule oder des nationalen Schulsystems geprägt werden. Auf allen drei Ebenen werden sodann Variablen einbezogen, um deren direkte und indirekte Zusammenhänge mit Schülerleistungen zu bestimmen. Eine große, repräsentative Stichprobe und differenzierte statistische Modelle gewährleisten eine genaue Absicherung der Ergebnisse, sowohl innerhalb als auch zwischen den teilnehmenden Staaten.

1 PISA kann Merkmale der Klassenebene wie beispielsweise Unterrichtspraktiken erheben und international vergleichend beschreiben. In Bedingungsanalysen kann die Klassenebene allerdings nur in einzelnen Teilnahmestaaten (z.B. Deutschland) im Rahmen nationaler Ergänzungen untersucht werden, da auf internationaler Ebene eine altersbasierte Stichprobenziehung vorgenommen wird, also Fünfzehnjährige quer über Jahrgänge und Klassen hinweg einbezogen werden.

Das Design der Fragebögen ist zunehmend professionalisiert worden und stellt inzwischen einen eigenständigen Kernbereich des internationalen PISA-Projekts dar. Während die beteiligten Wissenschaftler im ersten Zyklus der PISA-Studien vornehmlich damit befasst waren, Tests und dazu passende Fragebögen für den jeweils im Vordergrund stehenden Leistungsbereich zu entwickeln (Lesen bei PISA 2000, Mathematik im Jahr 2003, Naturwissenschaften in 2006), wird im derzeitigen zweiten Durchgang (PISA 2009, 2012, 2015) an einem integrierten Design zur Erfassung von Ausgangsbedingungen, Prozessmerkmalen und erzieherischen Wirkungen gearbeitet, das in aktuellen Theorien und Methoden der empirischen Bildungsforschung verankert ist. Hierzu wurden gesonderte Expertengruppen² eingesetzt, die entsprechende Rahmenkonzepte formulieren sollten. Neben dem Fragebogen-Design dient das jeweilige Rahmenkonzept auch der Vergewisserung über Fragestellungen, Aussagemöglichkeiten und Grenzen der Studien. Insofern sind gerade diese Konzepte wichtige Dokumente der Selbstreflexion der an PISA beteiligten Wissenschaftler.

Ziel des hier vorliegenden Aufsatzes ist es, diese Selbstreflexion für die deutsche Diskussion zugänglich zu machen. Die nachfolgenden Abschnitte beruhen wesentlich auf dem internationalen Rahmenkonzept für PISA 2012, das von den Autoren unter Beteiligung der internationalen Expertengruppe verfasst wurde (Klieme, Backhoff-Escudero et al., 2010). Im vorliegenden Aufsatz gehen wir jedoch vertiefend auf Forschungsliteratur ein und integrieren bereits Elemente des Rahmenkonzeptes für PISA 2015.

Nachfolgend (Abschnitt 2) wird zunächst am Beispiel PISA 2012 das Design der Studien dargestellt. Im Vordergrund stehen die in den Schüler-, Eltern- und Schulfragebogen ermittelten Merkmale, wobei auch erläutert werden soll, auf welcher theoretischen Basis diese Merkmale ausgewählt und operationalisiert werden. Anschließend (Abschnitt 3) werden aus der Nutzerperspektive, d.h. vor allem aus der Sicht von Bildungspolitik und -administration, Funktionen und Ziele der PISA-Studien erläutert. Es wird deutlich, dass PISA in der Tat an erster Stelle dem Bildungsmonitoring dient. Das internationale Rahmenkonzept geht aber auch auf die wissenschaftlichen Aussagemöglichkeiten und Grenzen der PISA-Studien ein, hier dargestellt in Abschnitt 4. Die Ausführungen zeigen insgesamt, inwieweit PISA anschlussfähig ist an die empirisch-pädagogische Forschung und andere Bereiche der Bildungsforschung.

2. Das PISA-Design und seine wissenschaftlichen Grundlagen

Um die verschiedenen Komponenten des Untersuchungsdesigns für internationale Leistungsvergleiche zu systematisieren, hat die *International Association for the Evaluation of Educational Achievement* bereits in den 1960er Jahren das sogenannte Kontext-Input-Prozess-Outcome-Modell (Purves, 1987) entwickelt, das sich ganz ähnlich in vielen

2 Die internationale Expertengruppe wurde bzw. wird bei PISA 2009 von Jaap Scheerens (Universität Twente, Niederlande), bei PISA 2012 und PISA 2015 von Eckhard Klieme (DIPF) geleitet.

aktuellen Arbeiten der empirischen Bildungsforschung findet. In diesem Modell werden Hintergrundvariablen und strukturelle Bedingungen (Inputs) unterschieden von Aktivitäten und Maßnahmen und deren Qualität (Prozesse) sowie den Leistungen und nicht-kognitiven (mit Fend: erzieherischen) Ergebnissen. Zum zweiten werden Handlungsebenen abgegrenzt, vom Bildungs- bzw. Schulsystem über die Ebene der Organisation (Einzelschule) und die Klassenebene zu Lernvoraussetzungen, -aktivitäten und Ergebnissen auf der Schülerebene. Die übergeordneten Ebenen werden als Kontext angesehen, der die jeweils untergeordneten Ebenen beeinflusst.

Eine aktuelle Version dieser im Englischen „CIPO-Modell“ genannten Matrix stellt Tabelle 1 dar. Diese Darstellung wurde dem internationalen Rahmenkonzept für PISA 2012 entnommen (Klieme, Backhoff-Escudero et al., 2010). Bei den in Tabelle 1 aufgeführten Merkmalen handelt es sich demnach um Konstrukte, die im Test- und Fragebogendesign der PISA-Erhebung 2012 international umgesetzt wurden. Gegenüber früheren Erhebungswellen konnte vor allem der Schülerfragebogen wesentlich erweitert werden, weil nun auch hier – wie schon seit PISA 2000 im Leistungstest – das Prinzip des sogenannten „Matrix-Designs“ eingesetzt wurde, bei dem die Schülerinnen und Schüler teilweise unterschiedliche, aber einander überlappende Mengen von Fragen bzw. Aufgaben bearbeiten. Dieses Prinzip erlaubte es im Jahr 2012, die Zahl der einbezogenen Konstrukte wesentlich zu vergrößern.

Die erste Spalte der Tabelle benennt die drei Ebenen, die in den PISA-Fragebögen explizit angesprochen sind: die Schüler-, die Klassen- und die Schulebene. Daneben werden die jeweiligen bei PISA erfassten Prozessmerkmale dargestellt sowie ggfs. Input- und Outcome-Merkmale. Prozessmerkmale sind alle Aspekte der Qualität und der Quantität des Lehrens und Lernens; darüber hinaus fallen professionelle Aktivitäten von Lehrerinnen und Lehrern und Schulleiterinnen und Schulleitern in diese Kategorie. Kognitive wie nicht-kognitive Ergebnisse (Outcomes) werden bei PISA ausschließlich auf der Ebene der Schülerinnen und Schüler durch Tests und Fragebögen erfasst; auf der Schul- und der Systemebene werden dann entsprechende Mittelwerte betrachtet. „Inputs“ werden bei Schülern und deren Eltern sowie auf der Schulebene (bei den Schulleitungen) erfragt, teilweise auch aus den Angaben der Schüler und Eltern aggregiert. Die Input-Variablen betreffen größtenteils den sozialen und familiären Hintergrund, aber auch Strukturmerkmale auf der Schulebene, zum Beispiel die Größe der Schule oder die Art der Finanzierung. Lehr- und Lernprozesse in der Klasse werden in den verschiedenen Fragebögen angesprochen und deshalb bei Tabelle 1 in einer gesonderten Zeile dargestellt; statistisch lässt die Stichprobenziehung der internationalen Studie (anders als die im vorliegenden Band ausgewertete deutsche Zusatzerhebung) jedoch nicht zu, einzelne Klassen zu identifizieren, sodass auch die Klassenmerkmale aggregiert auf der Schul- bzw. Systemebene ausgewertet werden.

Schülerinnen und Schüler	Input Geschlecht, Alter, sozio-ökonomischer Status, Migrationshintergrund Klassenstufe, Bildungskarriere, Klassenwiederholungen, Noten kulturelle Ressourcen, familiäre Lernumwelt und Unterstützung Erfahrung, Einstellungen und Fähigkeiten bzgl. Informations- und Kommunikationstechnologie Offenheit, Ausdauer und Engagement beim Problemlösen, Problemlösestile	Prozesse Lernzeit (inklusive Hausaufgaben und Nachhilfe) Teilnahme an außerunterrichtlichen Aktivitäten Unpünktlichkeit, Schulschwänzen schulisches Engagement schulbezogene und fachspezifische subjektive Normen, wahrgenommene Kontrolle des Lernprozesses, Attributions-Stil Akkulturationsstrategien	Outcomes Kompetenzen in Mathematik, Lesen, Naturwissenschaften, Problemlösen metakognitive Strategien Einstellung zum Fach, fachspezifische Motivation, fachspezifische Lernintentionen, fachspezifisches Selbstkonzept und Selbstwirksamkeit allgemeine schulbezogene Einstellungen, allgemeine Lernmotivation, Bildungsaspirationen
Klassen	Prozesse Unterrichtszeit, Leistungsgruppierungen, Leistungsbeurteilung und Feedback implementiertes Curriculum, Aufgaben, lernbezogene Aktivitäten, Nutzung von Lernmaterial Unterrichtsqualität: Disziplin/ Klassenklima, Unterstützung, Strukturierung, Schülerorientierung, kognitive Aktivierung		
Schulen	Input Schulform, Schulgröße, soziale und ethnische Komposition, Größe des Standorts Status (öffentlich versus privat), personelle und materielle Ausstattung, Wettbewerb mit anderen Schulen, mittlere Klassengröße Elternengagement, Leistungsdruck der Eltern	Prozesse Lehrerbildung/ Lehrerfortbildung Aufnahmekriterien der Schule, Bildungsgänge, Kursangebot/Curriculum, Leistungs differenzierung Lernzeit in den jeweiligen Fächern/ intendiertes Curriculum, Angebot an extracurricularen Aktivitäten, Angebote der Berufsberatung und -information geteilte Normen, Führungsstil, Arbeitsmoral der Lehrkräfte und Lehrerkooperation Schulentwicklung, Schulevaluation, Rechenschaftslegung, Schulautonomie, externe Kooperationen Schulklima, Schüler-Lehrer-Beziehungen, unterstützende Lernumgebung, Regelungen zum Umgang mit Schulschwänzen	

Tab. 1: Einordnung bei PISA 2012 erfasster Merkmale in eine Taxonomie der Ergebnisse („Outcomes“) von Bildungsprozessen und ihrer Bedingungen auf verschiedenen Ebenen des Bildungssystems

Auf die kognitiven Ergebnisvariablen, also die Leistungstests, die überwiegend im Zentrum der PISA-Rezeption stehen, soll an dieser Stelle nicht im Detail eingegangen werden (vgl. dazu etwa Klieme, Artelt et al., 2010). Aus erziehungswissenschaftlicher Sicht ist allerdings bemerkenswert, dass das Kompetenzkonzept der Studie letztlich auf die Pädagogische Anthropologie von Heinrich Roth (1971) zurückgeführt werden kann. Ausgehend von Mündigkeit als oberstem Ziel einer der Emanzipation verpflichteten Erziehung führt Roth den Kompetenzbegriff ein, um Mündigkeit auszudifferenzieren, und zwar in einem dreifachen Sinn, nämlich als Selbstkompetenz, Sachkompetenz und Sozialkompetenz (Roth, 1971, S. 180). Die in PISA erhobenen Grunddimensionen der Lese-, Mathematik- und Naturwissenschaft-Kompetenz sind als Sachkompetenzen im Sinne von Roth zu verstehen. Darüber hinaus erfasst PISA auch breitere kognitive Fähigkeiten, wie zum Beispiel Lernstrategien und Problemlösekompetenz. Ferner werden auch Aspekte der Selbstkompetenz sowie teilweise der Sozialkompetenz in den Kontextfragebögen berücksichtigt, so etwa Engagement, gemeinsame Werte und Überzeugungen, Motivation zu lernen und zu kooperieren und die Fähigkeit zur Selbstregulation eigener Lernprozesse. Die Bezüge zur Diskussion um Kompetenzen, wie sie Heinrich Roth in die Erziehungswissenschaft eingeführt hat, sind nicht zufällig: Über Publikationen von Fend und Weinert fanden sie Eingang in das OECD-Projekt „Definition and Selection of Competences“ sowie schließlich in die PISA-Studie (vgl. Klieme & Hartig, 2007).

Das CIPO-Modell ist zunächst nichts anderes als eine Taxonomie, die hilft, Konstrukte bzw. (auf der Ebene der Operationalisierungen) Messvariablen zu systematisieren. Häufig wird das Modell aber auch dazu genutzt, das Zusammenspiel von Faktoren auf verschiedenen Ebenen zu beschreiben. Man findet in solchen Darstellungen oft allerhand Pfeile, die andeuten sollen, dass „höhere“ auf „niedrigere“ Ebenen sowie Inputs auf Prozesse einwirken, die wiederum zu Outcomes führen. Dass eine solche einfache „Produktionsfunktion“, wie sie etwa Walberg (1986) in seinem einflussreichen Aufsatz darstellte, zu einfach ist, um die komplizierten Zusammenhänge, Wirkungen und Wechselwirkungen zu verstehen, ist aus erziehungswissenschaftlicher Sicht offensichtlich.

Schon die Unterscheidung zwischen Input, Prozess und Outcomes ist keineswegs eindeutig. Manche Prozessmerkmale (z.B. Lernstrategien) können ebenso gut als Inputs oder als Outcomes behandelt werden, abhängig von der theoretischen Perspektive, dem Forschungsdesign oder praktischen Überlegungen. In einem dynamischen Modell der Schuleffektivität (vgl. Creemers & Kyriakides, 2008) können zudem Outcomes als Inputs für die weitere Entwicklung angesehen werden. Angst vor Mathematik kann zum Beispiel als ein (negatives) Outcome schulischer Bildung angesehen werden, aber auch als ein Inputmerkmal, das einen Effekt auf weitere Prozess- und Outcome-Variablen, wie etwa Hausaufgabenaktivitäten, hat. Diese Überlegungen zeigen, dass das CIPO-Modell zwar ein nützliches Ordnungsschema ist, aber keine ausreichende theoretische Basis für systematische Wirkungsanalysen.

Um die Konstrukte und Messgrößen, die in das Design eingehen, auszuwählen und die daraufhin erhobenen Daten angemessen zu analysieren, nutzt PISA Wissen aus der

aktuellen Bildungsforschung. Zwei Forschungsfelder sind dabei von besonderer Bedeutung: die Schulwirkungsforschung sowie fachspezifische Lehr-Lern-Forschung.

Das zentrale Ziel der Schulwirkungsforschung ist es, „Faktoren im Bereich des Unterrichts, des Curriculums und der Lernumgebung auf verschiedenen Ebenen, wie der Klassen- und der Schulebene, [zu identifizieren], die – unter Berücksichtigung von Hintergrundmerkmalen wie der Fähigkeit, dem sozioökonomischen Status und der vorherigen Leistung – direkt oder indirekt Unterschiede zwischen Schülern bezüglich der Outcomes erklären können“ (Creemers & Kyriakides, 2008, S. 12; Übersetzung durch die Autoren). Im Sinne kognitiver Modelle des Lernens und Lehrens werden diese Faktoren nicht als direkt wirksame Maßnahmen gesehen, sondern als soziale Interaktionen und Lerngelegenheiten, die Schülerinnen und Schüler – ausgehend von ihren individuellen Fähigkeiten und Vorkenntnissen – durch eigene Lernaktivitäten für den Kompetenzerwerb nutzen. Die Schulwirkungsforschung kann mittlerweile auf einen beträchtlichen Wissensschatz zurückgreifen (z.B. Creemers & Kyriakides, 2008; Scheerens, 2000; Teddlie & Reynolds, 2000). Es ist vor allem das Verdienst von Jaap Scheerens, dieses Wissen systematisch für das Design der PISA-Fragebögen erschlossen zu haben.

Da PISA fachliche Kompetenzen beschreiben und erklären will, muss zudem auf fachspezifische Lehr-Lern-Forschung zurückgegriffen werden. Das Rahmenkonzept für PISA 2012 bezieht sich auf moderne Konzeptionen für den Mathematikunterricht wie „Lernen im Kontext“ (Schoenfeld, 2006) und „Diskussions-basierten Unterricht“ (Stein, Engle, Smith & Hughes, 2008), mit mathematischem Problemlösen und Modellieren (Blum & Leiss, 2005) als Kernaktivitäten. Die „derzeitige Vision eines wirksamen Mathematikunterrichts“, wie Stein und Koautoren (2008, S. 315) es nennen, sieht folgendermaßen aus: „Den Schülern werden vermehrt realistische und komplexe mathematische Probleme vorgelegt; sie unterstützen sich wechselseitig beim Lösen der Problemstellungen und diskutieren schließlich ihre Strategien und Lösungen in der Klasse unter Anleitung des Lehrers“ (Übersetzung durch die Autoren). Um diese Vision des Mathematikunterrichts abbilden, aber auch vom herkömmlichen Lösen standardisierter Aufgaben abgrenzen zu können, verwendet PISA 2012 gleich drei parallele, einander ergänzende Ansätze: Anhand exemplarischer Aufgabenstellungen werden Schüler und Schülerinnen gefragt, welche Lerngelegenheiten (opportunities to learn; Schmidt & Maier, 2009) sie aus ihrem Unterricht kennen. Zum zweiten werden Basisdimensionen der Unterrichtsqualität im Sinne von Klieme, Pauli und Reusser (2009) erfasst, und zum dritten wird die Häufigkeit der Anwendung lehrerzentrierter, schülerorientierter und individuell fördernder Unterrichtspraktiken ermittelt. Damit macht PISA 2012 einen innovativen, breiten Anlauf zu einer international vergleichenden Unterrichtsforschung, die in den Rahmen der Schulleistungsstudie mit Schwerpunkt Mathematik eingebaut ist.

3. Produkte der PISA-Studie im Rahmen des Bildungsmonitorings

Aus Sicht von Bildungspolitik und -administration kann PISA als ein Element der sogenannten neuen Steuerung der öffentlichen Verwaltung (New Public Management) be-

griffen werden, die Transparenz, Evaluation und Rechenschaftslegung eine große Bedeutung zumisst, bei einem gleichzeitigen Anstieg von Autonomie und Dezentralisierung (z.B. Böttcher, Bos, Döbert & Holtappels, 2008). Eine dauerhafte, datengestützte Beobachtung der Entwicklung von Bildungssystemen, wie sie PISA vornimmt, kann für die Evaluation und Rechenschaftslegung auf Systemebene genutzt werden. Die OECD selbst nutzt die Ergebnisse in ihren „Indikatoren für Bildungssysteme“. Deren Veröffentlichung regt öffentliche Debatten an, orientiert Bildungspolitik und liefert Grundlagen für Entscheidungen über Maßnahmen und Strategien in Bildungssystemen.

Dementsprechend entscheiden politische Fragestellungen über die Auswahl der Indikatoren. Bildungspolitik beschäftigt sich vor allem mit Ressourcen, Strukturen und Regeln für Prozessabläufe in Schulsystemen (z.B. Bildungsgänge, Curricula und Standards, personelle und materielle Ausstattung), mit den Bildungswegen und -ergebnissen (etwa dem Niveau und Profil der Schülerleistungen) und schließlich auch mit Fragen der Chancengleichheit (z.B. wie Ressourcen und Zugangschancen verteilt werden). Entsprechend reflektieren die in PISA gebildeten Indikatoren diese grundlegenden Themen.

Eine Vielzahl von Indikatoren, die auf den PISA-Fragebögen basieren, werden in den verschiedenen Ausgaben des OECD-Berichts „Bildung auf einen Blick“ dargestellt (vgl. zuletzt OECD, 2011). Neben der reinen Dokumentation von Leistungsergebnissen waren dies in der Vergangenheit:

- Der Zusammenhang von Geschlecht, sozialer Herkunft und Migrationshintergrund mit der Schülerleistung (nahezu jährlich, von 2002-A10 bis 2011-A5);
- Einstellungen von Schülern wie Bildungsaspiration, Lernmotivation und Interesse, Selbstregulation, Leseverhalten und Zugehörigkeitsgefühl zur Schule sowie deren Zusammenhänge mit individuellen Voraussetzungen einerseits, Leistungen andererseits (2003-A9 & A10, 2004-A8, 2007-A4 & A5, 2011-A6);
- Profile von Schülerinnen und Schülern, die herausragende Leistungen erbringen, unter anderem deren Einstellungen und Motivation (2009-A4 & A5); analog: Profile von Risikoschülern (2006-A6);
- Zusammenhänge zwischen Outcomes und Ressourcen (2007 & 2008-B7), vor allem der Klassengröße (2008-D2);
- Institutionelle Differenzierung im Bildungssystem – verstanden als horizontale und vertikale Differenzierung – und deren Zusammenhang mit Schülerleistungen (2002-A7, 2006-A7);
- Vergleich zwischen öffentlichen und privaten Schulen (2005-D5);
- Outcomes von berufsbezogenen, verglichen mit allgemeinen Bildungsgängen (2007 & 2008-C1);
- Die Nutzung von Evaluation und Leistungsbeurteilung in Bildungssystemen (2008-D5);
- Schulklima und Schulqualität aus der Perspektive von Eltern (2008-A5) bzw. Schulleitungen (2002-D5);
- Chancenverteilung bezüglich des Zugangs zu qualifizierter Bildung (2007-A4 &

A7, 2008-A3 & A7, 2011-D6) oder technologiegestützten Lerngelegenheiten (2002-D3 & D4) in Abhängigkeit vom sozialen und familiären Hintergrund der Schülerinnen und Schüler.

Ob auf PISA beruhende Indikatoren für Politik und Bildungsadministration informativ und nützlich sind, hängt nicht zuletzt von der Breite und Aussagefähigkeit des Designs ab. Ein Mittel, die Angemessenheit des Designs und der darauf beruhenden Indikatoren zu prüfen, ist ein Abgleich mit der aktuellen Forschungsliteratur, etwa dem „Handbook of Education Policy Research“ (Sykes, Schneider & Plank, 2009). In seinen 76 Kapiteln werden verschiedenste Aspekte dessen behandelt, was man im Deutschen wohl als „Steuerung des Bildungswesens“ übersetzen würde – jedoch ohne den mit dem Steuerungsbegriff konnotierten Anspruch auf Durchgriff „von oben“. Der Abschnitt des Buches, der sich mit „Policy implications of resources, management and organization“ befasst, beinhaltet acht „steuerungsrelevante“ Themen wie Klassengröße, Chancengerechtigkeit, Schulwahl und Ressourcen; für alle diese Themen kann PISA dank des Fragebogen-Designs Kennwerte liefern. Dasselbe gilt für den Abschnitt „Teaching and learning policy“ mit Themen wie Lerngelegenheiten und Qualifikation der Lehrkräfte sowie den Abschnitt „Educational access and differentiation“ zu Bildungswegen – einschliesslich früher Bildung –, ethnischer Diskriminierung oder dem „Schattensystem“ von Nachhilfe und privaten Zusatzschulen. Praktisch alle Themen des Handbuchs können durch PISA-Indikatoren langfristig, international vergleichend beobachtet werden, mit Ausnahme der Fragen zu Entscheidungsprozessen in der Bildungspolitik, z.B. zur Rolle der Lehrgewerkschaften.

Die besondere Chance eines auf PISA aufbauenden Bildungsmonitorings liegt im internationalen Vergleich der Indikatoren, der allerdings voraussetzt, dass die relevanten Indikatoren sich tatsächlich in international äquivalenter Weise messen lassen (Hui & Triandis, 1985; van de Vijver & Leung, 1997; van de Vijver, van Hemert & Poortinga, 2008). Die Vergleichbarkeit von Indikatoren soll durch das Einbeziehen internationaler Experten, durch nationale Adaptationen und durch einen Feldtest gewährleistet werden. Für Leistungstests und Fragebogenskalen lässt sich deren Vergleichbarkeit auch empirisch überprüfen mithilfe von Analysen differenzieller Itemfunktionen und Mehrgruppenfaktorenanalysen. Hier geht es vor allem um den Einfluss kulturspezifischer Werte, Normen und mentaler Schemata auf die Bedeutung von Konstrukten, die Interpretation einzelner Fragen sowie das Antwortverhalten (van de Vijver & Leung, 1997). Generell wird in PISA eine hohe interkulturelle Äquivalenz sowohl der Tests als auch der Fragebogenskalen erzielt (vgl. OECD, 2005). Gleichzeitig deuten aber verschiedene Sekundäranalysen darauf hin, dass ein Teil der internationalen Varianz in den Mittelwerten der Fragebogenskalen durch methodische Artefakte wie etwa kulturbedingte Unterschiede im Antwortverhalten erzeugt wird (vgl. van de Gaer, Grisay, Schulz & Gebhardt, 2012). Aufschlussreichere Informationen für das Monitoring nicht-kognitiver Prozesse und Outcomes können insofern Trendanalysen innerhalb eines Landes erbringen, da diese von kulturellen Einflüssen auf die Definition und Messung der Konstrukte nicht betroffen sind. Deshalb ist es für die Nutzung der PISA-Daten zum Systemmoni-

toring von zentraler Bedeutung, sicherzustellen, dass die Indikatoren über verschiedene Erhebungswellen und Zyklen hinweg vergleichbar sind. Gleichzeitig muss es aber auch möglich sein, neue Indikatoren aufzunehmen und zu etablieren, um Aktualität und Innovation zu gewährleisten.

4. Die Bedeutung von PISA für das Verständnis der Wirksamkeit von Bildungssystemen

Indikatoren lenken die öffentliche Aufmerksamkeit auf erfolgreiche und weniger erfolgreiche Sektoren von Bildungssystemen. Sie informieren darüber, welche Ziele schon erreicht worden sind und wo sich nach wie vor Herausforderungen für die Bildungspolitik ergeben. In diesem Sinne dienen sie als Grundlage für bildungspolitische Entscheidungen und das Setzen von Prioritäten. Die bei PISA eingesetzten Hintergrundfragebögen wurden aber stets so gestaltet, dass sie auch Hinweise zu Zusammenhängen der Outcomes untereinander sowie mit Input- und Prozessmerkmalen geben können. Grundlegende Voraussetzung hierfür ist, dass die wichtigsten Input- und Prozessvariablen auf allen drei Analyseebenen erhoben werden. Statistische Modelle zur Analyse der Mehrebenenaten können helfen, die komplexen Beziehungen zwischen familiären und individuellen Bedingungen, staatlichen Regelungen, pädagogischer Praxis sowie schließlich Lernprozessen und -ergebnissen zu erschließen. Dabei gilt es zu berücksichtigen, dass die Schul- und Unterrichtsmerkmale nicht für alle Schülerinnen und Schüler innerhalb einer Schule gleich effektiv sein müssen und sich die Effekte außerdem je nach lokalem Kontext oder Schule unterscheiden können. Deshalb versucht moderne Schulforschung, auch Interaktionen zwischen Input- und Prozessmerkmalen zu berücksichtigen und differenzielle Zusammenhänge zu untersuchen (Creemers & Kyriakides, 2008). Solche komplexen Modellvorstellungen lassen sich bei PISA aufgrund der großen Stichprobe gut prüfen. Sofern Informationen über Ressourcen und Kosten erhoben werden, kann PISA schließlich dazu beitragen, Effizienz zu verstehen, also Effektivität in Relation zu den eingesetzten Ressourcen.

Als internationale Studie ermöglicht PISA neben einfachen Vergleichen der Ausprägung von Inputs, Prozessen und Outcomes der beteiligten Systeme auch die Analyse von interkulturellen Ähnlichkeiten und Unterschieden in der Struktur der Merkmale und ihrer Zusammenhänge. Damit lässt sich die Generalisierbarkeit der gängigen Effektivitätsmodelle prüfen. Dies ist umso bedeutsamer, als die meisten veröffentlichten Befunde der Bildungsforschung immer noch aus westlichen Kulturen stammen. PISA kann mit seinen inzwischen etwa 70 beteiligten Staaten dazu beitragen, den einseitig westlichen Blick auf Bildung und Schule zu überwinden. Beispielsweise diskutieren Coll, Dahsah & Faikhamta (2010) anhand von PISA-2006-Daten, dass Kontextfaktoren in westlichen und asiatischen Ländern unterschiedliche Bedeutung zukommt. Sofern internationale Varianz beobachtet wird, können zudem Mehrebenenanalysen helfen, Zusammenhänge mit Systemmerkmalen und kulturellen Werten zu untersuchen. Beispiele hierfür sind die Rolle von Strukturmerkmalen des Bildungssystems bei der Erklärung

von Bildungsaspirationen (Dupriez, Monseur, van Campenhoudt & Lafontaine, 2012) oder Moderationseffekte kultureller Wertorientierungen auf den Zusammenhang zwischen schulischen Prozessfaktoren, wie der Lehrerunterstützung, und Schülerleistungen (Vieluf, 2012).

Die Breite der wissenschaftlichen Erkenntnisse, die mit PISA verbunden sind, kann exemplarisch anhand von Vorträgen bei der ersten PISA-Forschungs-Konferenz (Prenzel, Kobarg, Schöps & Rönnebeck, 2013)³ dargestellt werden. Während einige der Studien auf international vergleichenden Analysen basieren, beschreiben andere nur Ergebnisse für einzelne Staaten.

- In Staaten mit niedrigem Leseengagement ist die Chancengleichheit geringer als in Staaten, in denen Schülerinnen und Schüler mehr lesen und mehr Freude am Lesen berichten (Baye, Monseur & Lafontaine, 2009).
- Bei der Förderung naturwissenschaftlicher Kompetenz und des naturwissenschaftlichen Interesses von Schülerinnen und Schülern scheint ein fokussiertes Muster naturwissenschaftlichen Lernens erfolgversprechender als eine breite Vielfalt naturwissenschaftlicher Aktivitäten (Kobarg, Seidel, Prenzel, McCrae & Walker, 2009).
- Das Interesse von Schülerinnen und Schülern an einer Karriere im Bereich der Naturwissenschaften hängt zusammen mit dem Ausmaß, in dem Schulen auf solche Karrieren vorbereiten (Lie & Kjaernsli, 2009).
- In Schweizer Schulen korrelieren Unterrichtszeit und Mathematikleistung signifikant – außer in Gymnasien (Angelone, Moser & Ramseier, 2009).
- Während japanische Schülerinnen und Schüler großes Interesse an entdeckendem Lernen zeigen, ist der Oberstufenunterricht in Japan nicht auf diese Interessen zugeschnitten. Dieses Ergebnis könnte möglicherweise zur Erklärung des vergleichsweise geringen Interesses an Naturwissenschaften in Japan beitragen (Yasushi, 2009).
- In Südkorea finden sich Zusammenhänge der Schülerleistungen mit den Ressourcen einer Schule sowie mit dem Ausmaß, in dem die Schule Schülerinnen und Schüler ermutigt, ihr volles Potential zu nutzen (Kaplan, 2009a).

Für die wissenschaftliche Grundlagenforschung von besonderer Bedeutung sind Erkenntnisse zur Validität zentraler Konstrukte, wie etwa die Arbeit von Marsh, Hau, Artelt, Baumert & Peschar (2006) zur interkulturell äquivalenten Erfassung von Aspekten der Selbstregulation bei PISA 2000 oder die Analysen von Drechsel, Carstensen und Prenzel (2011) zur Dimensionalität des Interesses an Naturwissenschaften, basierend auf PISA 2006.

Wenngleich PISA einen Beitrag dazu leisten kann, Wissen über Bildungspolitik und -praxis zu kreieren, müssen dabei wichtige Grenzen beachtet werden. Eine Einschränkung ist, dass PISA den Ertrag von Bildungsprozessen misst, also Fähigkeiten und Kompetenzen, die seit der frühen Kindheit beim Durchlaufen verschiedener Schulen

3 Zusammenfassungen aller Präsentationen, die im Folgenden zitiert werden, sind erhältlich auf <http://www.pisaresconf09.org>

bis zum Alter von 15 Jahren akkumuliert worden sind. Damit kann PISA nicht ermitteln, wie viel Lernen tatsächlich in der von den Schülerinnen und Schülern mit 15 Jahren besuchten Schule der Sekundarstufe stattgefunden hat. Nur wenn das Leistungsniveau zum Zeitpunkt des Eintritts in die aktuelle Schule erfasst worden wäre, könnte der Fortschritt (*value-added*) in der Leistung mit den bildungsbezogenen Erfahrungen in der spezifischen Schule in Bezug gesetzt werden. Solche Messungen der Leistung zu einem Zeitpunkt vor der PISA-Erhebung stehen jedoch nicht zur Verfügung. Auch der Einfluss von Lehrerinnen und Lehrern auf Schülerleistungen kann anhand von PISA nicht beurteilt werden – zumindest nicht mit dem Design, das während des ersten Jahrzehnts auf internationaler Ebene implementiert worden ist (zu den deutschen Erweiterungen vgl. Klieme, Artelt et al., 2010). Der Grund hierfür ist, dass lediglich eine Zufallsstichprobe Fünfzehnjähriger in jeder Schule gezogen wird, anstelle von vollständigen Klassen, was die Erhebung von Informationen über Unterrichtsstrategien und Unterrichtsqualität in einer bestimmten Klasse ausschließt. Lediglich Deutschland untersucht regelmäßig auch vollständige Klassen (vgl. die Beiträge des vorliegenden Bandes zu Befunden aus der Jahrgangsstufe 9). Schließlich muss berücksichtigt werden, dass etwa in einem Fünftel der Staaten eine Mehrheit der Schülerinnen und Schüler erst kurz vor der Erhebung auf eine weiterführende Schule gewechselt ist. Dies erschwert einmal mehr Schlüsse über Schuleffekte in den entsprechenden Staaten.

Weiterhin muss bei der Interpretation der Ergebnisse berücksichtigt werden, dass es generell ausgesprochen schwierig ist, auf der Basis einer Survey-Studie mit Leistungsdaten zu belegen, dass eine bestimmte Maßnahme oder Praxis einen direkten Einfluss oder eine tatsächliche Wirkung auf Schülerleistungen hat (Gustafsson, 2007; Kaplan, 2009b). Wird etwa beobachtet, dass Schulen höhere Schülerleistungen aufzuweisen haben, wenn sie ihre Schulevaluationsdaten veröffentlichen, so erlaubt diese querschnittlich ermittelte Korrelation keine Schlüsse darüber, ob die Veröffentlichung zu einer höheren Leistung beiträgt oder ob es vielmehr vor allem erfolgreiche Schulen sind, die ihre Ergebnisse veröffentlichen. Wenn die OECD aus dem beobachteten Zusammenhang die politische Schlussfolgerung zieht, externes Monitoring könne den Leistungsstand einer Schule verbessern (OECD, 2007, S. 276), liegt also eine wissenschaftlich unzulässige Interpretation vor. Auch wenn komplexe statistische Verfahren zur Analyse von Kausaleffekten eingesetzt werden (z.B. *Matching*-Verfahren; siehe dazu Heckman, 2005; Kaplan, 2009b; Schneider, Carnoy, Kilpatrick, Schmidt & Shavelson, 2007), würde eine solche Interpretation voraussetzen, dass alle Variablen, die potentiell mit der Auswahl des Treatments sowie mit dem interessierenden Ergebnis zusammenhängen, ebenfalls erfasst worden sind. Selbst im Rahmen einer so groß angelegten Untersuchung wie PISA ist dies schwerlich umzusetzen.

Der Wert von PISA ergibt sich nicht zuletzt aus dem kontinuierlichen Zusammenspiel mit anderen Untersuchungen. Auf der einen Seite kann mithilfe von Längsschnittstudien, experimentellen Untersuchungen und Interventionsstudien überprüft werden, inwieweit in PISA beobachtete Korrelationen tatsächlich auf einem Kausalzusammenhang beruhen. Auf der anderen Seite können auf der Grundlage der aktuellen Forschungsliteratur jeweils jene Variablen als Indikatoren für das Bildungsmonitoring aus-

gewählt werden, deren Bedeutung für Bildungsqualität und -effizienz in vorherigen Untersuchungen, im Idealfall in Längsschnitt- oder Interventionsstudien, belegt worden ist.

Das Konzept der Wirksamkeit wird hier in dem technischen Sinne verwendet, der Kausalitätsanalysen in der ökonomischen und psychologischen Forschung zugrunde liegt, ausgehend von den wissenschaftstheoretischen und statistischen Ansätzen bei Donald Rubin (1974). In der Perspektive eines akteurszentrierten Institutionalismus, wie ihn – aufbauend auf internationaler soziologischer Forschung – Helmut Fend (2006) in seiner Neuen Theorie der Schule eingeführt hat, wären solche Ursache-Wirkungs-Gefüge dahingehend zu hinterfragen, wie Regelungen und Vorgaben, einschließlich systematischer „Interventionen“, auf allen beteiligten Ebenen rekontextualisiert werden. Hierzu kann eine quantitative Studie wie PISA nur begrenzt beitragen. Immerhin lässt sich im Rahmen des mehrperspektivischen Fragebogen-Designs systematisch untersuchen, wie bestimmte Phänomene wie etwa Unterrichtspraktiken, Kooperation und Partizipation von unterschiedlichen Akteursgruppen an einer Schule wahrgenommen werden. In den Planungen und Vorstudien für PISA 2015 werden solche Ansätze derzeit erprobt.

Auf lange Sicht wird die Verfügbarkeit von Trend-Daten der größte Nutzen der PISA-Datenbank sein. Querschnittsstudien tragen nur wenig zum Verständnis der Entstehung von Outcomes bei, weil es praktisch unmöglich ist, alle relevanten Variablen abzudecken – inklusive Informationen über die Schülerleistung bei Eintritt in die Sekundarstufe. Demgegenüber erlauben Veränderungen in Outcomes auf der Schul- oder Schülerebene, aber auch solche auf der Ebene von Staaten (vgl. Gustafsson, 2007; Hanushek & Wößmann, 2010), substantiellere Schlussfolgerungen – vor allem wenn sie mit Veränderungen von Inputs und Prozessen in Beziehung gesetzt werden können. Um dieses Ziel zu erreichen, muss PISA einen Variablensatz definieren, der über die Erhebungszyklen konstant gehalten wird. Es wäre außerdem aufschlussreich, wenn im Rahmen von PISA zunehmend auch längsschnittliche Designs auf Schul- und Schülerebene implementiert würden, wie es in Deutschland mehrfach geschehen ist (vgl. etwa Bischof et al., in diesem Band).

5. Fazit

Der vorliegende Beitrag hatte zum Ziel, die Motive, Aussagemöglichkeiten und Grenzen der PISA-Studie sowie ihren praktischen und theoretischen Hintergrund zusammenfassend darzustellen. Die Darlegungen spiegeln den internationalen Arbeitsstand bei der Entwicklung der Rahmenkonzeption, die den Kontext der Leistungsmessungen abdeckt, also die relevanten Input-, Prozess- und Outcome-Merkmale auf Schüler-, Klassen- und Schulebene.

Eine zentrale Aufgabe von PISA ist es, Indikatoren für das international vergleichende Monitoring von Bildungssystemen bereitzustellen. Gleichzeitig dient die Studie aber dem Aufbau von Wissen darüber, *warum* spezifische Merkmalskonstellationen in verschiedenen Bildungssystemen vorgefunden werden. Die Auswahl der Merkmale

ist von politischen Interessen geleitet, beruht aber andererseits auf Erkenntnissen der empirisch-pädagogischen Forschung und benachbarter Disziplinen, insbesondere der Schuleffektivitätsforschung und der Lehr-Lern-Forschung in den Testdomänen (Lesen, Mathematik, Naturwissenschaften und Problemlösen).

Bildungsmonitoring und Wirkungsanalyse: Diese beiden Modi der Gestaltung und Interpretation charakterisieren internationale Schulstudien schon seit mehr als einem halben Jahrhundert. Ihre Koexistenz führt immer wieder zu Missverständnissen und Kurzschlüssen, sowohl bei Autoren als auch bei Kritikern des Paradigmas. Um Überinterpretationen und unangemessene Bedeutungszuschreibungen zu vermeiden, ist es von zentraler Bedeutung, die Aussagemöglichkeiten und Einschränkungen der Studien zu bestimmen. Beispielsweise jene, dass sich Wirkungsfragen aufgrund einer querschnittlichen Survey-Studie kaum beantworten lassen.

Im vorliegenden Aufsatz wurde versucht, die Chancen und Grenzen von PISA als Instrument des Bildungsmonitorings und als wissenschaftliche Datenquelle zu beleuchten unter Hinweis auf Literatur, die entweder in PISA aufgegriffen wird oder umgekehrt PISA-Daten für Forschungszwecke verwendet. Der Aufsatz kann daher als Dokumentation der Selbstreflexion von Wissenschaftlern angesehen werden, die bei PISA engagiert sind. Eine externe Bewertung der Studie, die in der deutschen Erziehungswissenschaft an prominenter Stelle rezipiert wurde, hat Münch (2009) vorgelegt. Er charakterisiert PISA als „Benchmarking der OECD-Mitgliedstaaten in der Erfüllung global definierter Bildungsstandards“ unter der „Leitidee der Bildung als Humankapital“ (S. 31 und 33) und bewertet den Erkenntnisgewinn als „eher bescheiden“ (S. 220). Es sei dem Leser überlassen, ein eigenes Fazit zu bilden.

Literatur

- Angelone, D., Moser, U., & Ramseier, E. (2009). *Instruction time and performance. Analyses of the importance of instruction time for competencies in Science and Mathematics based on Swiss PISA 2006 Data* (paper presented at the PISA Research Conference, Kiel, Germany, 14-16 September 2009).
http://www.pisaresconf09.org/user_uploads/files/context/room3/Angelone_Moser_Ramseier.pdf
 [20.12.2012].
- Baumert, J., Lehmann, R., Lehrke, M., Schmitz, B., Clausen, M., Hosenfeld, I., Köller, O., & Neubrand, J. (1997). *TIMSS – Mathematisch-naturwissenschaftlicher Unterricht im internationalen Vergleich. Deskriptive Befunde*. Opladen: Leske + Budrich.
- Baye, A., Monseur, C., & Lafontaine, D. (2009). *Institutional and socioeconomic factors influencing the reading engagement* (paper presented at the PISA Research Conference, Kiel, Germany, 14-16 September 2009).
http://www.pisaresconf09.org/user_uploads/files/content/room1/Baye_Monseur_Lafontaine_2.pdf
 [20.12.2012].
- Blömeke, S., Kaiser, G., & Lehmann, R. (Hrsg.) (2010). *TEDS-M 2008 – Professionelle Kompetenz und Lerngelegenheiten angehender Mathematiklehrkräfte für die Sekundarstufe I im internationalen Vergleich*. Münster: Waxmann.
- Blum, W., & Leiss, D. (2005). How Do Students and Teachers Deal with Mathematical Modeling Problems? The Example “Sugarloaf”. *ICTMA 12 Proceedings*, 222-231.

- Böttcher, W., Bos, W., Döbert, H., & Holtappels, H. G. (Hrsg.) (2008). *Bildungsmonitoring und Bildungscontrolling in nationaler und internationaler Perspektive*. Münster: Waxmann.
- Coll, R. K., Dahsah, C., & Faikhamta, C. (2010). The Influence of Educational Context on Science Learning: A Cross-National Analysis of PISA. *Research in Science & Technological Education*, 28(1), 3-24.
- Creemers, B. P. M., & Kyriakides, L. (2008). *The Dynamics of Educational Effectiveness: A Contribution to Policy, Practice, and Theory in Contemporary Schools*. London/New York: Routledge.
- Drechsel, B., Carstensen, C., & Prenzel, M. (2011). The Role of Content and Context in PISA Interest Scales: A Study of the Embedded Interest Items in the PISA 2006 Science Assessment. *International Journal of Science Education*, 33(1), 73-95.
- Dupriez, V., Monseur, C., van Campenhoudt, M., & Lafontaine, D. (2012). Social Inequalities of Post-secondary Educational Aspirations: Influence of Social Background, School Composition and Institutional Context. *European Educational Research Journal*, 11(4), 504-519.
- Fend, H. (2006). *Neue Theorie der Schule*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Gustafsson, J.-E. (2007). Understanding Causal Influences on Educational Achievement through Analysis of Differences over Time within Countries. In T. Loveless (Hrsg.), *Lessons Learned: What International Assessments Tell Us about Math Achievement* (S. 37-63). Washington: The Brookings Institution.
- Hanushek, E. A., & Wößmann, L. (2010). *The Economics of International Differences in Educational Achievement* (IZA Discussion Paper No. 4925). Bonn: IZA.
- Heckman, J. J. (2005). The Scientific Model of Causality. *Sociological Methodology*, 35, 1-97.
- Hui, C. H., & Triandis, H. C. (1985). Measurement in Cross-cultural Psychology: A Review and Comparison of Strategies. *Journal of Cross-Cultural Psychology*, 16, 131-152.
- Kaplan, D. (2009a). *Advances in Multilevel Latent Variable Models for PISA Data* (paper presented at the PISA Research Conference, Kiel, 14-16 September 2009). http://www.pisaresconf09.org/user_uploads/files/methodology/room1/Kaplan.pdf [20.12.2012].
- Kaplan, D. (2009b). Causal Inference in Non-experimental Educational Policy Research. In D. N. Plank, W. E. Schmidt & G. Sykes (Hrsg.), *AERA Handbook on Education Policy Research* (S. 139-153). Washington: American Educational Research Association.
- Keeves, J. P., & Lietz, P. (2011). The Relationship of IEA to Some Developments in Educational Research Methodology and Measurement during the Years from 1962 to 1992. In C. Papanastasiou, T. Plomp & E. C. Papanastasiou (Hrsg.), *IEA 1958-2008: 50 Years of Experiences and Memories* (S. 217-252). Nicosia: Cultural Center of the Kykkos Monastery.
- Klieme, E. (2011). Bildung unter undemokratischem Druck? Anmerkungen zur Kritik der PISA-Studie. In S. Aufenanger, S. F. Hamburger, L. Ludwig & R. Tippelt (Hrsg.), *Bildung in der Demokratie. Beiträge zum 22. Kongress der DGfE*, Band 2 (S. 289-302). Opladen/Farmington Hills: Barbara Budrich.
- Klieme, E., Artelt, C., Hartig, J., Jude, N., Köller, O., Prenzel, M., Schneider, W., & Stanat, P. (2010). *PISA 2009. Bilanz nach einem Jahrzehnt*. Münster: Waxmann.
- Klieme, E., Backhoff-Escudero, E., Blum, W., Buckley, J., Hong, Y.-Y., Kaplan, D., Levin, H. M., & Vieluf, S. (2010). *Designing PISA as a Sustainable Data Base for Educational Policy and Research. The PISA 2012 Context Questionnaire Framework*. Paris: OECD.
- Klieme, E., & Hartig, J. (2007). Kompetenzkonzepte in den Sozialwissenschaften und im empirischen Diskurs. In M. Prenzel, I. Gogolin & H.-H. Krüger (Hrsg.), *Kompetenzdiagnostik* (Zeitschrift für Erziehungswissenschaft, Sonderheft 8, S. 11-29).
- Klieme, E., Pauli, C., & Reusser, K. (2009). The Pythagoras Study: Investigating Effects of Teaching and Learning in Swiss and German Classrooms. In T. Janik & T. Seidel (Hrsg.), *The Power of Video Studies in Investigating Teaching and Learning in the Classroom* (S. 137-160). Münster: Waxmann.

- Kobarg, M., Seidel, T., Prenzel, M., McCrae, B., & Walker, M. (2009). *Patterns of Science Teaching and Learning in an International Comparison* (paper presented at the PISA Research Conference, Kiel, Germany, 14-16 September 2009).
http://www.pisaresconf09.org/user_uploads/files/context/room3/Kobarg_etal.pdf [20.12.2012].
- Lie, S., & Kjaernsli, M. (2009). *Predictors for Students' Choice of Science Careers* (paper presented at the PISA Research Conference, Kiel, Germany, 14-16 September 2009).
http://www.pisaresconf09.org/user_uploads/files/context/room3/Lie_Kjaernsli.pdf [20.12.2012].
- Marsh, H. W., Hau, K.-T., Artelt, C., Baumert, J., & Peschar, J. L. (2006). OECD's Brief Self-report Measure of Educational Psychology's Most Useful Affective Constructs: Cross-cultural, Psychometric Comparisons Across 25 Countries. *International Journal of Testing*, 6, 311-360.
- Münch, R. (2009). *Globale Eliten, lokale Autoritäten: Bildung und Wissenschaft unter dem Regime von PISA*, McKinsey & Co. Frankfurt a.M.: Suhrkamp.
- OECD (Hrsg.) (2005). *PISA 2003 Technical Report*. Paris: OECD.
- OECD (Hrsg.) (2007). *PISA 2006 – Science Competencies for Tomorrow's World*. Paris: OECD.
- OECD (Hrsg.) (2011). *Education at a Glance*. Paris: OECD.
- Ozga, J. (2008). Governing Knowledge: Research Steering and Research Quality. *European Educational Research Journal*, 7(3), 261-272.
- Ozga, J. (2012). Assessing PISA. *European Educational Research Journal*, 11(2), 166-171.
- Prenzel, M., Kobarg, M., Schöps, K., & Rönnebeck, S. (2013). *Research on PISA. Research Outcomes of the PISA Research Conference 2009*. New York: Springer.
- Purves, A. C. (1987). The Evolution of the IEA: A Memoir. *Comparative Education Review*, 31(1), 10-28.
- Roth, H. (1971). *Pädagogische Anthropologie. Band 2: Entwicklung und Erziehung*. Hannover: Schroedel.
- Rubin, D. B. (1974). Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies. *Journal of Educational Psychology*, 66, 688-701.
- Scheerens, J. (2000). *Improving School Effectiveness* (Fundamentals of Educational Planning Series, IIEP, Vol. 68). Paris: UNESCO.
- Schmidt, W. H., & Maier, A. (2009). Opportunity to Learn. In G. Sykes, B. L. Schneider & D. N. Plank (Hrsg.), *Handbook on Education Policy Research* (S. 541-549). New York: Routledge.
- Schneider, B. L., Carnoy, M., Kilpatrick, J., Schmidt, W. H., & Shavelson, R. J. (2007). *Estimating Causal Effects: Using Experimental and Observational Designs*. Washington: American Educational Research Association.
- Schoenfeld, A. (2006). Mathematics Teaching and Learning. In P. A. Alexander & I. H. Winne (Hrsg.), *Second Handbook of Educational Psychology* (S. 479-510). London: Routledge.
- Stein, M. K., Engle, R. A., Smith, M. S., & Hughes, E. K. (2008). Orchestrating Productive Mathematical Discussions: Five Practices for Helping Teachers Move Beyond Show and Tell. *Mathematical Thinking and Learning*, 10(4), 313-340.
- Sykes, G., Schneider, B. L., & Plank, D. N. (2009). *Handbook of Education Policy Research*. New York: Routledge.
- Teddlie, C., & Reynolds, D. (Hrsg.) (2000). *The International Handbook of School Effectiveness Research*. New York: Routledge.
- van de Gaer, E., Grisay, A., Schulz, W., & Gebhardt, E. (2012). The Reference Group Effect: An Explanation of the Paradoxical Relationship Between Academic Achievement and Self-Confidence Across Countries. *Journal of Cross-Cultural Psychology*, 43(8), 1205-1228.
- van de Vijver, F. J. R., & Leung, K. (1997). *Methods and Data Analysis for Cross-cultural Research*. Newbury Park: Sage.
- van de Vijver, F. J. R., van Hemert, D. A., & Poortinga, Y. H. (Hrsg.) (2008). *Multilevel Analysis of Individuals and Cultures*. New York: Erlbaum.

- Vieluf, S. (2012). *Links between Student Achievement and Teacher Support in Different Education Systems* (Vortrag bei der Arbeitstagung Methoden des interkulturellen Vergleichs, unveröffentlichtes Manuskript). Frankfurt a.M.: DIPF.
- Walberg, H. J. (1986). Syntheses of Research on Teaching. In M. C. Wittrock (Hrsg.), *Handbook of Research on Teaching* (3. Aufl., S. 214-229). New York: Macmillan.
- Walker, D. F. (1976). Toward Comprehension of Curricular Realities. *Review of Research in Education*, 4, 268-308.
- Yasushi, O. (2009). *Comparison of Attitudes toward Science between Grade 9 and 10 Japanese Students by Using the PISA Questions and Its Implications on Science Teaching in Japan* (paper presented at the PISA Research Conference, Kiel, Germany, 14-16 September 2009). http://www.pisaresconf09.org/user_uploads/files/content/room2/OGURA_Yasushi.pdf [20.12.2012].

Anschrift des Autors/der Autorin

Prof. Dr. Eckhard Klieme, Deutsches Institut für Internationale Pädagogische Forschung (DIPF), Schloßstraße 29, 60486 Frankfurt am Main, Deutschland
E-Mail: klieme@dipf.de

Dipl.-Psych. Svenja Vieluf, Deutsches Institut für Internationale Pädagogische Forschung (DIPF), Schloßstraße 29, 60486 Frankfurt am Main, Deutschland
E-Mail: vieluf@dipf.de



Doris Bühler-Niederberger

**Lebensphase
Kindheit**

Theoretische Ansätze, Akteure
und Handlungsräume

Grundlagentexte Soziologie, hrsg. von
M. Diewald und K. Hurrelmann.

2011, 256 Seiten, broschiert,

€ 19,95

ISBN 978-3-7799-1488-4

Eine systematische und umfassende Einführung in die aktuelle sozialwissenschaftliche Kindheitsforschung, ihre theoretischen Grundlagen und Konzepte, ihre Datenquellen und ihren empirischen Ertrag stand bisher aus und soll nun mit diesem Band geleistet werden. Die verschiedenen theoretischen Ansätze werden vorgestellt und auf ihre Stärken und Schwächen hin gesichtet. Ein differenziertes und strukturiertes Bild wird vermittelt vom Aufwachsen in Deutschland und der Varietät von Kindheiten, die man in einer internationalen und historischen Sicht konstatieren kann.

Aus dem Inhalt:

Die lange, behütete Kindheit und ihre gesellschaftliche Konstruktion

Aufwachsen in Deutschland – lange, behütete Kindheit als soziale Realität und normatives Muster

Kindheiten in der Welt – zwischen lokalen Bedingungen und globalen Einflüssen

Geschichte der Kindheit – lange, behütete Kindheit als Element sozialer Ordnung

Soziologische Theorien zu Kindern und Kindheit

Theorien der Sozialisation

„Neue“ Soziologie der Kindheit

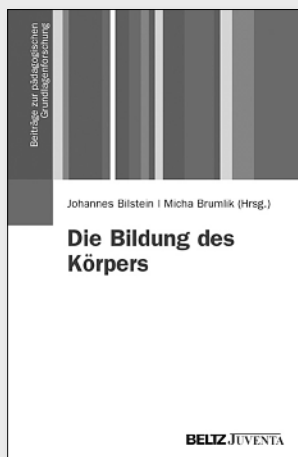
Synthese und Ausblick – „Generationales Ordnen“ als Kernkonzept eines Programms der Kindheitsforschung

www.juventa.de

BELTZ JUVENTA



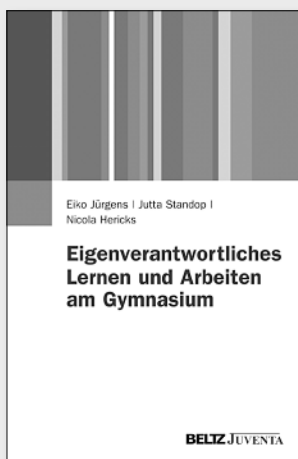
7., korrigierte Auflage 2012
326 Seiten, broschiert, € 19,95
ISBN 978-3-7799-2178-3



Reihe: Beiträge zur pädagogischen
Grundlagenforschung
2013, 336 Seiten, broschiert, € 32,95
ISBN 978-3-7799-1270-5



2012, 332 Seiten, broschiert,
€ 24,95
ISBN 978-3-7799-2818-8



2012, 264 Seiten, broschiert
€ 32,95
ISBN 978-3-7799-2812-6

www.juventa.de

BELTZ JUVENTA